

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 09:24:29 ; Search time 54 Seconds
(without alignments)
88,950 Million cell updates/sec

Title: US-09-998-491-7

Perfect score: 83

Sequence: 1 MVQSMRHRKAELEKA 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*

- 1: Geneseq19808:*
- 2: Geneseq19908:*
- 3: Geneseq20008:*
- 4: Geneseq20018:*
- 5: Geneseq20028:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|--------------------|
| 1 | 83 | 100.0 | 17 | 6 | Abb99610 Peptide d |
| 2 | 77 | 92.8 | 19 | 6 | Abb99631 Amino aci |
| 3 | 45 | 54.2 | 10 | 6 | Abb99630 Amino aci |
| 4 | 44 | 53.0 | 325 | 4 | Abb71638 Drosophil |
| 5 | 43 | 51.8 | 500 | 7 | Adc37415 Nuclear f |
| 6 | 43 | 51.8 | 536 | 7 | Adc37417 Nuclear f |
| 7 | 43 | 51.8 | 536 | 7 | Adc37419 Nuclear f |
| 8 | 43 | 51.8 | 559 | 4 | Abg21539 Novel hum |
| 9 | 43 | 51.8 | 720 | 4 | Abb58376 Drosophil |
| 10 | 42.5 | 51.2 | 305 | 4 | Abb61456 Human TAN |
| 11 | 42.5 | 51.2 | 329 | 3 | Abb15545 Human TAN |
| 12 | 42.5 | 51.2 | 329 | 4 | Abb78822 Human pro |
| 13 | 42.5 | 51.2 | 329 | 4 | Abb93503 Human pro |
| 14 | 42.5 | 51.2 | 329 | 4 | Abb61418 Human TAN |
| 15 | 42.5 | 51.2 | 329 | 7 | AdD19036 Human dis |
| 16 | 42.5 | 51.2 | 337 | 4 | Abb11797 Human sec |
| 17 | 42.5 | 51.2 | 337 | 4 | Abb79806 Human pro |
| 18 | 42 | 50.6 | 258 | 4 | Aag74570 Human col |
| 19 | 42 | 50.6 | 447 | 7 | AdB63969 Human pro |
| 20 | 42 | 50.6 | 685 | 5 | Abb77429 Human tan |
| 21 | 42 | 50.6 | 885 | 4 | Abb61554 Drosophil |
| 22 | 42 | 50.6 | 1372 | 2 | Aaw56473 Protein w |
| 23 | 42 | 50.6 | 1379 | 7 | AdB79765 Rat Rho k |
| 24 | 42 | 50.6 | 1388 | 2 | Aaw56475 Protein w |
| 25 | 42 | 50.6 | 1388 | 5 | Abb57354 Mouse isc |

ALIGNMENTS

RESULT 1
ABB99610
ID ABB99610 standard; peptide; 17 AA.
XX
AC ABB99610;
XX
DT 28-MAR-2003 (first entry)
XX
DE Peptide derived from human amyloid precursor protein (APP).
XX
KW Amyloid precursor protein; APP; protein derivative;
KW neurodegenerative disease; Alzheimer's disease; cognitive enhancer.
XX
OS Synthetic.
OS Homo sapiens.
XX
FN WQ200283729-A2.
XX
PD 24-OCT-2002.
XX
PF 17-APR-2002; 2002WO-GB001769.
XX
PR 18-APR-2001; 2001GB-00009558.
PR 17-AUG-2001; 2001GS-00020084.
PR 30-NOV-2001; 2001US-00998491.
PR 28-MAR-2002; 2002GB-00007387.
XX
PA (UYOP-) UNIV OPEN.
XX
PI Mileusnic R, Rose SPR;
XX
XX WPI; 2003-111814/10.
XX
DR Derivatives of polypeptides, useful for treating neurodegenerative
XX disease e.g. Alzheimer's disease, comprises one functional amino acid
XX residue or derivative protected by a protective group.
XX
PS Disclosure; Page 1; 85pp; English.
XX
CC The present sequence is derived from amyloid precursor protein (APP).
CC Derivatives of the invention are based on APP sequences. The
CC specification describes a derivative of a polypeptide in which at least
CC one functional group of at least one amino acid residue or derivative is
CC protected by a protective group. This derivative is of the formula given
CC in ABB99625. The derivative is useful in medicine and in the preparation
CC of a medicament for use in the treatment of a neurodegenerative disease
CC e.g. Alzheimer's disease. It is also useful as a cognitive enhancer

```
SQ Sequence 17 AA;
Query Match 100.0%; Score 83; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQSMRERHKAELREKA 17
   ||||| |||||
DB 1 MVQSMRERHKAELREKA 17

RESULT 2
ABB99631
ID ABB99631 standard; peptide; 19 AA.
XX AC ABB99631;
XX DT 28-MAR-2003 (first entry)
XX DE Amino acid sequence of a peptide derivative.
XX KW Amyloid precursor protein; APP; protein derivative;
XX KW neurodegenerative disease; Alzheimer's disease; cognitive enhancer.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1
XX FT /note= "0-25 natural or synthetic amino acids or
XX FT derivatives thereof"
XX FT Misc-difference 8
XX FT /note= "a natural or synthetic amino acid or derivative
XX FT thereof, with at least on functional group of the amino
XX FT acid being protected by a protective group"
XX FT Misc-difference 19
XX FT /note= "0-21 natural or synthetic amino acids or
XX FT derivatives thereof"
XX PN WO200283729-A2.
XX PR 17-APR-2002; 2002WO-GB001769.
XX PD 24-OCT-2002.
XX PF 17-APR-2002; 2002WO-GB001769.
XX PR 18-APR-2001; 2001GB-00009558.
XX PR 17-AUG-2001; 2001GB-00020084.
XX PR 30-NOV-2001; 2001US-00998491.
XX PR 28-MAR-2002; 2002GB-00007387.
XX PA (UYOP-) UNIV OPEN.
XX XX Mileusmic R, Rose SPR;
XX DR WPI; 2003-111814/10.
XX PT Derivatives of polypeptides, useful for treating neurodegenerative
XX PT disease e.g. Alzheimer's disease, comprises one functional amino acid
XX PT residue or derivative protected by a protective group.
XX PS Claim 19; Page; 85pp; English.
XX CC The present sequence represents a derivative of the invention.
XX CC Derivatives of the invention may be based on amyloid precursor protein
XX CC (APP) sequences. The specification describes a derivative of a
XX CC polypeptide in which at least one functional group of at least one amino
XX CC acid residue or derivative is protected by a protective group. This
XX CC derivative is of the formula given in ABB99625. The derivative is useful
XX CC in medicine and in the preparation of a medicament for use in the
XX CC treatment of a neurodegenerative disease e.g. Alzheimer's disease. It is
XX CC also useful as a cognitive enhancer. note: this sequence does not appear
XX CC in the specification; it was created using information provided in the
XX CC claims section

SQ Sequence 19 AA;
Query Match 92.8%; Score 77; DB 6; Length 19;
Best Local Similarity 94.1%; Pred. No. 1.1e-05;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVQSMRERHKAELREKA 17
   ||||| ||||| |||||
DB 2 MVQSMRERHKAELREKA 18

RESULT 3
ABB99630
ID ABB99630 standard; peptide; 10 AA.
XX AC ABB99630;
XX DT 28-MAR-2003 (first entry)
XX DE Amino acid sequence of a peptide derivative fragment.
XX KW Amyloid precursor protein; APP; protein derivative;
XX KW neurodegenerative disease; Alzheimer's disease; cognitive enhancer.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Misc-difference 10
XX FT /note= "comprises 0-21 natural or synthetic amino acids
XX FT or derivatives thereof"
XX PN WO200283729-A2.
XX PD 24-OCT-2002.
XX PF 17-APR-2002; 2002WO-GB001769.
XX PR 18-APR-2001; 2001GB-00009558.
XX PR 17-AUG-2001; 2001GB-00020084.
XX PR 30-NOV-2001; 2001US-00998491.
XX PR 28-MAR-2002; 2002GB-00007387.
XX PA (UYOP-) UNIV OPEN.
XX XX Mileusmic R, Rose SPR;
XX DR WPI; 2003-111814/10.
XX PT Derivatives of polypeptides, useful for treating neurodegenerative
XX PT disease e.g. Alzheimer's disease, comprises one functional amino acid
XX PT residue or derivative protected by a protective group.
XX PS Claim 19; Page 53; 85pp; English.
XX CC The present sequence represents a fragment of a derivative of the
XX CC invention. The present peptide represents the first Xaa of ABB99625.
XX CC Derivatives of the invention may be based on amyloid precursor protein
XX CC (APP) sequences. The specification describes a derivative of a
XX CC polypeptide in which at least one functional group of at least one amino
XX CC acid residue or derivative is protected by a protective group. This
XX CC derivative is of the formula given in ABB99625. The derivative is useful
XX CC in medicine and in the preparation of a medicament for use in the
XX CC treatment of a neurodegenerative disease e.g. Alzheimer's disease. It is
XX CC also useful as a cognitive enhancer
XX CC Sequence 10 AA;
Query Match 54.2%; Score 45; DB 6; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.88;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 HKAELEKA 17
   |||||
```


XX 03-DEC-2001; 2001JP-00368692.
 PR 05-DEC-2001; 2001US-0335829P.
 PR 03-OCT-2002; 2002JP-00291302.
 PR 04-OCT-2002; 2002US-0415769P.
 XX (ASAH) ASAH KASEI KK.
 XX Matsuda A, Muramatsu S;
 XX WPI; 2003-505282/47.
 DR N-PSDB; ADC37416.
 XX New purified protein that activates nuclear factor kappa B (NF-kappaB),
 PT useful for treating inflammation, autoimmune diseases, cancers,
 PT infectious diseases, bone diseases, AIDS, neurodegenerative diseases or
 PT ischemic disorders.
 XX Claim 1; SEQ ID NO 250; 938pp; English.
 XX The present invention relates to novel proteins and their coding
 CC sequences (ADC37168-ADC37455), which activate nuclear factor kappa B (NF-
 CC kappaB). The proteins and their coding sequences are useful for treating
 CC a disease associated with NF-kappaB activation, such as inflammation,
 CC autoimmune diseases, cancers, infectious diseases, bone diseases, AIDS,
 CC neurodegenerative diseases, or ischaemic disorders.
 XX Sequence 536 AA;
 SQ
 Query Match 51.8%; Score 43; DB 7; Length 536;
 Best Local Similarity 47.1%; Pred. No. 1.4e+02;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MVQSMREPHKAEUREKA 17
 DB 312 MAQKEKEHKEUREMA 328
 |||:|:|:|:|
 |||:|:|:|:|
 RESULT 7
 ADC37419
 ID ADC37419 standard; protein; 536 AA.
 XX
 AC ADC37419;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Nuclear factor kappa B (NF-kappaB) activating protein, SEQ ID 252.
 XX
 KW Nuclear factor kappa B; NF-kappaB; inflammation; autoimmune disease;
 KW cancer; infectious disease; bone disease; AIDS;
 KW neurodegenerative disease; ischaemic disorder; Antiinflammatory;
 KW Immunomodulator; Cytostatic; Antimicrobial; Osteopathic; Anti-HIV;
 KW Neuroprotective; Nootropic; Cardiant; Gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 XX WC2003048202-A2.
 XX
 PD 12-JUN-2003.
 XX
 PF 03-DEC-2002; 2002WO-JP012644.
 XX
 PR 03-DEC-2001; 2001JP-00368692.
 PR 05-DEC-2001; 2001US-0335829P.
 PR 03-OCT-2002; 2002JP-00291302.
 PR 04-OCT-2002; 2002US-0415769P.
 XX (ASAH) ASAH KASEI KK.
 XX Matsuda A, Muramatsu S;
 XX WPI; 2003-505282/47.
 DR N-PSDB; ADC37416.

XX New purified protein that activates nuclear factor kappa B (NF-kappaB),
 PT useful for treating inflammation, autoimmune diseases, cancers,
 PT infectious diseases, bone diseases, AIDS, neurodegenerative diseases or
 PT ischemic disorders.
 XX Claim 1; SEQ ID NO 252; 938pp; English.
 XX The present invention relates to novel proteins and their coding
 CC sequences (ADC37168-ADC37455), which activate nuclear factor kappa B (NF-
 CC kappaB). The proteins and their coding sequences are useful for treating
 CC a disease associated with NF-kappaB activation, such as inflammation,
 CC autoimmune diseases, cancers, infectious diseases, bone diseases, AIDS,
 CC neurodegenerative diseases, or ischaemic disorders.
 XX Sequence 536 AA;
 SQ
 Query Match 51.8%; Score 43; DB 7; Length 536;
 Best Local Similarity 47.1%; Pred. No. 1.4e+02;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MVQSMREPHKAEUREKA 17
 DB 312 MAQKEKEHKEUREMA 328
 |||:|:|:|:|
 |||:|:|:|:|
 RESULT 8
 ABG21539
 ID ABG21539 standard; protein; 559 AA.
 XX
 AC ABG21539;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #21530.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 XX WC200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS85726.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 51898; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 559 AA;

Query Match 51.8%; Score 43; DB 4; Length 559;
 Best Local Similarity 47.1%; Pred. No. 1.5e+02;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 MVQSMRERHKAELREKA 17
 DB 322 MAQKEKEHEKLEMA 338

RESULT 9

ABE58376
 ID ABE58376 standard; protein; 720 AA.

XX
 AC ABE58376;

XX
 DT 26-MAR-2002 (first entry)

XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 1920.

XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX
 OS Drosophila melanogaster.

XX
 FN WO200171042-A2.

XX
 PD 27-SEP-2001.

XX
 PF 23-MAR-2001; 2001WO-US009231.

XX
 PR 23-MAR-2000; 2000US-0191637P.

XX
 PR 11-JUL-2000; 2000US-00614150.

XX
 PA (PEKE) PE CORP NY.

XX
 PI Venter JC, Adams M, Li PWD, Myers EW;

XX
 XX WPI; 2001-656860/75.

XX
 DR N-PSDB; ABL02479.

XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

XX
 PS Disclosure; SEQ ID NO 1920; 21pp + Sequence Listing; English.

XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABE57737-
 CC ABE72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 720 AA;

Query Match 51.8%; Score 43; DB 4; Length 720;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 MRRERHKAELREK 16

DB 430 LEDRHOALLREK 441

RESULT 10

AAB61456
 ID AAB61456 standard; protein; 305 AA.

XX
 AC AAB61456;

XX
 DT 04-APR-2001 (first entry)

XX
 DE Human TANGO 246 domain.

XX
 KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung; autoimmune;
 KW allergy; cardiovascular; brain; degenerative; placental; pancreatic;
 KW skeletal; muscle.

XX
 OS Homo sapiens.

XX
 FN WO200100672-A1.

XX
 PD 04-JAN-2001.

XX
 PF 29-JUN-2000; 2000WO-US018184.

XX
 PR 29-JUN-1999; 99US-00342687.

XX
 PA (WILL-) MILLENNIUM PHARM INC.

XX
 PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;

XX
 WPI; 2001-050127/06.

XX
 PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases
 PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
 PT disorders (e.g. jaundice).

XX
 PS Disclosure; Page 235-236; 262pp; English.

XX
 CC The present invention relates to cDNAs encoding TANGO 244, TANGO 246,
 CC TANGO 275, TANGO 300 and MANGO 245 proteins. The nucleic acids, proteins
 CC and protein modulators are useful for treating colonic disorders,
 CC inflammatory diseases, tumors, renal disorders, liver disorders, lung
 CC disorders, autoimmune diseases, allergic diseases, cardiovascular
 CC diseases, brain disorders, degenerative diseases placental, pancreatic,
 CC skeletal and muscle disorders

XX
 SQ Sequence 305 AA;

Query Match 51.2%; Score 42.5; DB 4; Length 305;
 Best Local Similarity 62.5%; Pred. No. 91;
 Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 3 QSMRERHKAEL-REKA 17

DB 238 QAMQELHVELEREKA 253

RESULT 11

AAB15545
 ID AAB15545 standard; protein; 329 AA.

XX
 AC AAB15545;

XX
 DT 28-FEB-2001 (first entry)

XX Human immune system molecule from Incyte clone 1352789.
 DE Anti-inflammatory; keratolytic; anti-HIV; anti-allergic; antianaemic;
 KW antiarteriosclerotic; antiasthmatic; antidiabetic; nephrotropic; cancer;
 KW antigout; dermatologic; antithyroid; virucide; hepatotropic; antibody;
 KW immunosuppressive; cytostatic; fungicide; protozoacide; antibacterial;
 KW gene therapy; diagnostic; immunological disorder; viral infection;
 KW bacterial infection; fungal infection; parasitic infection; immunogen.
 XX Homo sapiens.
 OS WO2000060080-A2.
 XX 12-OCT-2000.
 XX 04-APR-2000; 2000WO-US009072.
 XX 05-APR-1999; 99US-0127852P.
 PR 05-MAY-1999; 99US-0132647P.
 XX (INCY-) INCYTE PHARM INC.
 PA Yue H, Lal P, Tang YT, Baughn MR, Azimzai Y, Lu DAM;
 PI WPI; 2000-665005/64.
 DR N-PSDB; AAA95784.
 XX New human immune system molecules 1-15 and polynucleotides encoding them
 FT useful for diagnosing, treating or preventing e.g. immunological
 PT disorders, infections, cell proliferative disorders, microbial
 PT infections.
 XX Claim 1; Page 82-83; 95pp; English.
 PS This sequence represents a human immune system molecule (IMOL) encoded by
 CC the cDNA isolated as clone 1352789 from the Incyte IATRTU02 library. The
 CC human IMOLs (AAB15536-B15550) and their encoding polynucleotides
 CC (AAA95775-A95789), and compositions comprising them are useful for the
 CC diagnosis, treatment or prevention of immunological disorders, infections
 CC and cell proliferative disorders, including cancer. The IMOL may be used
 CC to treat or prevent disorders associated with decreased expression or
 CC activity of IMOLs, such as immunological disorders (e.g. inflammation,
 CC actinic keratosis, AIDS, Addison's disease), haematopoietic cancer,
 CC infections caused by virus (e.g. adenovirus, parvovirus, coronavirus),
 CC bacteria (e.g. Staphylococcus, Streptococcus, Shigella), fungi (e.g.
 CC Aspergillus, Blastomycetes), parasites (e.g. Plasmodium, Trypanosoma,
 CC intestinal protozoa), cell proliferative disorders (e.g. actinic
 CC keratosis, arteriosclerosis, bursitis), and cancers (e.g. leukemia,
 CC melanoma, sarcoma). The peptides are also useful as immunogens for the
 CC development of antibodies that specifically recognize these peptides.
 CC The polynucleotides may be used to detect and quantify gene expression in
 CC biopsied tissues in which expression of IMOL may be correlated with the
 CC disease, as targets in a microarray, to detect differences in gene
 CC sequences among normal, carrier and affected individuals, and for
 CC screening libraries of compounds in drug screening techniques. Antibodies
 CC which specifically bind to IMOL may be used for the diagnosis of
 CC disorders characterized by expression of IMOL, or in assays to monitor
 CC patients being treated with IMOL or agonists, antagonists, or inhibitors
 CC of IMOL
 XX Sequence 329 AA;
 SQ Query Match 51.2%; Score 42.5; DB 3; Length 329;
 Best Local Similarity 62.5%; Pred. No. 99;
 Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
 QY 3 QSMRERHKAEL-REKA 17
 DB 238 QANQELHVELEREKA 253
 RESULT 12
 ID AAB93503 standard; protein; 329 AA.

AAW78822
 ID AAW78822 standard; protein; 329 AA.
 XX
 AC AAW78822;
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 1484.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 OS WO200157190-A2.
 XX 09-AUG-2001.
 XX 05-FEB-2001; 2001WO-US004098.
 XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZN;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 DR N-PSDB; AAK51955.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 FT in diagnosis and gene therapy.
 XX Claim 20; Page 3765; 6221pp; English.
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX Sequence 329 AA;
 SQ Query Match 51.2%; Score 42.5; DB 4; Length 329;
 Best Local Similarity 62.5%; Pred. No. 99;
 Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
 QY 3 QSMRERHKAEL-REKA 17
 DB 238 QANQELHVELEREKA 253
 RESULT 13
 ID AAB93503 standard; protein; 329 AA.

XX AAB93503;
AC
XX
DT 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:12824.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
XX EP1074617-A2.
XX
XX PD 07-FEB-2001.
XX
XX PF 28-JUL-2000; 2000EP-00116126.
XX
XX PR 29-JUL-1999; 99JP-00248036.
XX
XX PR 27-AUG-1999; 99JP-00300253.
XX
XX PR 11-JAN-2000; 2000JP-00118776.
XX
XX PR 02-MAY-2000; 2000JP-00183767.
XX
XX PR 09-JUN-2000; 2000JP-00241899.
XX
XX PA (HELI-) HELIX RES INST.
XX
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
XX PS Claim 8; SEQ ID NO 12824; 2537pp + Sequence Listing; English.
XX
XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
XX
XX CC (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the
XX
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the
XX
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
XX SQ Sequence 329 AA;
XX
XX Query Match 51.2%; Score 42.5; DB 4; Length 329;
XX
XX Best Local Similarity 62.5%; Pred. No. 99;
XX
XX Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
XX
XX QY 3 QSMRERHKAEL-REKA 17
XX
XX Db 238 QAMQELHVELEREKA 253
XX
XX RESULT 14
XX
XX AAB61418
XX
XX Query Match 51.2%; Score 42.5; DB 4; Length 329;
XX
XX Best Local Similarity 62.5%; Pred. No. 99;
XX
XX Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
XX
XX QY 3 QSMRERHKAEL-REKA 17
XX
XX Db 238 QAMQELHVELEREKA 253
XX
XX RESULT 14
XX
XX AAB61418

AD AAB61418 standard; protein; 329 AA.
XX
XX AC AAB61418;
XX
XX DT 04-APR-2001 (first entry)
XX
XX DE Human TANGO 246 protein.
XX
XX XX TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung; autoimmune;
XX
XX KW allergy; cardiovascular; brain; degenerative; placental; pancreatic;
XX
XX KW skeletal; muscle.
XX
XX OS Homo sapiens.
XX
XX XX WO200100672-A1.
XX
XX PN 04-JAN-2001.
XX
XX PD 29-JUN-2000; 2000WO-US018184.
XX
XX PF 29-JUN-1999; 99US-00342687.
XX
XX PR (MILL-) MILLENNIUM PHARM INC.
XX
XX PA Holtzman DA, Barnes TM, Fraser CC, Sharp JD;
XX
XX PI WPI; 2001-050127/06.
XX
XX DR Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
XX
XX PT MANGO 245 proteins, useful in the treatment of inflammatory diseases
XX
XX PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
XX
XX PT disorders (e.g. jaundice).
XX
XX XX Claim 1; Fig 5; 262pp; English.
XX
XX CC The present invention relates to cDNAs encoding TANGO 244, TANGO 246,
XX
XX CC TANGO 275, TANGO 300 and MANGO 245 proteins. The nucleic acids, proteins
XX
XX CC and protein modulators are useful for treating colonic disorders,
XX
XX CC inflammatory diseases, tumors, renal disorders, liver disorders, lung
XX
XX CC disorders, autoimmune diseases, allergic diseases, cardiovascular
XX
XX CC diseases, brain disorders, degenerative diseases placental, pancreatic,
XX
XX CC skeletal and muscle disorders
XX
XX SQ Sequence 329 AA;
XX
XX Query Match 51.2%; Score 42.5; DB 4; Length 329;
XX
XX Best Local Similarity 62.5%; Pred. No. 99;
XX
XX Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
XX
XX QY 3 QSMRERHKAEL-REKA 17
XX
XX Db 238 QAMQELHVELEREKA 253
XX
XX RESULT 15
XX
XX ADD19036
XX
XX ID ADD19036 standard; protein; 329 AA.
XX
XX AC ADD19036;
XX
XX XX 15-JAN-2004 (first entry)
XX
XX DE Human disease related protein SeqID525.
XX
XX KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
XX
XX KW antiarteriosclerotic; vulnary; gene therapy;
XX
XX KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
XX
XX KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
XX
XX KW glucose transportation; catecholamine synthesis; iron transport;
XX
XX KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
XX
XX KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
XX
XX KW inflammatory condition; wound healing.
XX

| | | |
|----|---|--------------------------------------|
| OS | Homo sapiens. | |
| XX | | |
| PN | WO2003018621-A2. | |
| XX | | |
| PD | 06-MAR-2003. | |
| XX | | |
| PF | 23-AUG-2002; 2002WO-GE003892. | |
| XX | | |
| PR | 23-AUG-2001; 2001GB-00020558. | |
| PR | 05-OCT-2001; 2001GB-00024037. | |
| PA | (OXFO-) OXFORD BIOMEDICA UK LTD. | |
| XX | | |
| PI | Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR; | |
| DR | WPI; 2003-290046/28. | |
| DR | N-PSDB; ADD19037. | |
| XX | | |
| PT | New substantially purified polypeptide, useful for diagnosing or treating | |
| PT | a hypoxia-regulated condition, such as cancer, ischemia, reperfusion | |
| PT | injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or | |
| PT | wound healing. | |
| XX | | |
| PS | Claim 1; SEQ ID NO 525; 424pp; English. | |
| XX | | |
| CC | This invention relates to novel human genes and gene product which are | |
| CC | implicated in certain disease states. Compounds which modulate the | |
| CC | proteins of the invention may have cytostatic, antiinflammatory, | |
| CC | ophthalmological, antiarteriosclerotic or vulnerary activities. The | |
| CC | sequences of the invention may be useful for gene therapy. The invention | |
| CC | may be useful for diagnosing or treating a hypoxia-regulated condition, | |
| CC | such as tumorigenesis, angiogenesis, apoptosis, inflammation, | |
| CC | erythropoiesis, or the biological response to hypoxia conditions | |
| CC | including processes such as glycolysis, gluconeogenesis, glucose | |
| CC | transportation, catecholamine synthesis, iron transport or nitric oxide | |
| CC | synthesis. The disease includes cancer, ischaemic conditions, reperfusion | |
| CC | injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis, of | |
| CC | inflammatory conditions or wound healing. The present sequence is that of | |
| CC | a disease related protein of the invention. | |
| XX | | |
| XX | | |
| QQ | Sequence 329 AA; | |
| | Query Match | 51.2%; Score 42.5; DB 7; Length 329; |
| | Best local Similarity | 62.5%; Pred. No. 99; |
| | Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1; | |
| QY | 3 QSMRRHKAEL-REXA 17 | |
| | : : : : : | |
| DB | 238 QAMQELHVELEREKA 253 | |
| | : : : : : | |
| | Search completed: March 9, 2004, 09:25:40 | |
| | Job time : 56 secs | |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 09:24:35 ; Search time 23 seconds
(without alignments)
38.158 Million cell updates/sec

Title: US-09-998-491-7

Perfect score: 83

Sequence: 1 MVQSMRERHKAELREKA 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2 6/ptodata/2/iaa/5A COMB.pep:*
- 2: /cgn2 6/ptodata/2/iaa/5B COMB.pep:*
- 3: /cgn2 6/ptodata/2/iaa/6A COMB.pep:*
- 4: /cgn2 6/ptodata/2/iaa/6B COMB.pep:*
- 5: /cgn2 6/ptodata/2/iaa/PTUS COMB.pep:*
- 6: /cgn2 6/ptodata/2/iaa/backfilesi.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 83 | 100.0 | 17 | 5 | PCT-US92-09070-17 |
| 2 | 42 | 50.6 | 119 | 3 | US-08-685-871-66 |
| 3 | 42 | 50.6 | 1388 | 2 | US-08-685-576-1 |
| 4 | 42 | 50.6 | 1388 | 2 | US-08-685-576-4 |
| 5 | 42 | 50.6 | 1388 | 4 | US-09-765-594-296 |
| 6 | 41 | 49.4 | 257 | 2 | US-08-715-204-6 |
| 7 | 41 | 49.4 | 257 | 3 | US-09-162-597-6 |
| 8 | 40 | 48.2 | 529 | 4 | US-08-887-534A-74 |
| 9 | 40 | 48.2 | 529 | 4 | US-09-527-431-74 |
| 10 | 39 | 47.0 | 144 | 4 | US-09-250-609-2 |
| 11 | 39 | 47.0 | 144 | 4 | US-09-250-611-2 |
| 12 | 39 | 47.0 | 202 | 4 | US-09-489-039A-11417 |
| 13 | 39 | 47.0 | 204 | 2 | US-08-715-204-1 |
| 14 | 39 | 47.0 | 204 | 2 | US-08-691-814B-10 |
| 15 | 39 | 47.0 | 204 | 3 | US-09-162-597-1 |
| 16 | 39 | 47.0 | 204 | 4 | US-09-250-609-4 |
| 17 | 39 | 47.0 | 204 | 4 | US-09-250-609-9 |
| 18 | 39 | 47.0 | 204 | 4 | US-09-250-611-4 |
| 19 | 39 | 47.0 | 204 | 4 | US-09-250-611-9 |
| 20 | 39 | 47.0 | 218 | 4 | US-09-489-039A-10326 |
| 21 | 39 | 47.0 | 266 | 4 | US-09-252-991A-26842 |
| 22 | 39 | 47.0 | 279 | 4 | US-09-252-991A-31984 |
| 23 | 39 | 47.0 | 384 | 4 | US-09-134-001C-3437 |
| 24 | 39 | 47.0 | 563 | 4 | US-09-134-001C-4800 |
| 25 | 38.5 | 46.4 | 168 | 2 | US-08-667-023-2 |
| 26 | 38.5 | 46.4 | 168 | 2 | US-08-713-825-3 |
| 27 | 38.5 | 46.4 | 168 | 3 | US-09-199-842-3 |

| | | | | | | |
|----|------|------|-----|---|----------------------|-------------------|
| 28 | 38.5 | 46.4 | 168 | 4 | US-09-460-532-3 | Sequence 3, Appli |
| 29 | 38.5 | 46.4 | 463 | 4 | US-09-489-039A-12860 | Sequence 12860, A |
| 30 | 38 | 45.8 | 286 | 1 | US-08-202-186-25 | Sequence 25, Appl |
| 31 | 38 | 45.8 | 286 | 2 | US-08-418-071-18 | Sequence 18, Appl |
| 32 | 38 | 45.8 | 287 | 4 | US-09-252-991A-30447 | Sequence 30447, A |
| 33 | 38 | 45.8 | 307 | 3 | US-09-049-672A-9 | Sequence 9, Appli |
| 34 | 38 | 45.8 | 359 | 4 | US-09-328-352-7295 | Sequence 7295, Ap |
| 35 | 38 | 45.8 | 633 | 4 | US-09-252-991A-23497 | Sequence 23497, A |
| 36 | 38 | 45.8 | 993 | 3 | US-09-060-410-4 | Sequence 4, Appli |
| 37 | 38 | 45.8 | 993 | 4 | US-09-723-458-4 | Sequence 4, Appli |
| 38 | 37.5 | 45.2 | 166 | 4 | US-09-252-991A-18518 | Sequence 18518, A |
| 39 | 37 | 44.6 | 185 | 2 | US-08-691-814B-12 | Sequence 12, Appl |
| 40 | 37 | 44.6 | 185 | 4 | US-09-250-609-11 | Sequence 11, Appl |
| 41 | 37 | 44.6 | 185 | 4 | US-09-250-611-11 | Sequence 11, Appl |
| 42 | 37 | 44.6 | 384 | 4 | US-09-252-991A-30256 | Sequence 30256, A |
| 43 | 37 | 44.6 | 412 | 4 | US-09-252-991A-22952 | Sequence 22952, A |
| 44 | 37 | 44.6 | 429 | 4 | US-08-795-876-44 | Sequence 44, Appl |
| 45 | 37 | 44.6 | 472 | 4 | US-09-252-991A-20558 | Sequence 20558, A |

ALIGNMENTS

RESULT 1
PCT-US92-09070-17
; Sequence 17, Application PC/TUS9209070
; GENERAL INFORMATION:
; APPLICANT: Saitoh, Tsunao [NM]
; TITLE OF INVENTION: SUBSTANCES HAVING THE GROWTH-PROMOTING
; EFFECT OF AMYLOID PRECURSOR PROTEIN
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive
; CITY: Newport Beach
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09070
; FILING DATE: 19921023
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: UC035.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
PCT-US92-09070-17

Query Match 100.0%; Score 83; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVQSMRERHKAELREKA 17
Db 1 MVQSMRERHKAELREKA 17

RESULT 2

US-08-685-871-66
 ; Sequence 66, Application US/08685871
 ; Patent No. 6013499
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAMURA, Shu
 ; APPLICANT: IWAMATSU, Akihiro
 ; TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
 ; NUMBER OF SEQUENCES: 68
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/685,871
 ; FILING DATE: 24-JUL-1996
 ; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 8-184102
 ; FILING DATE: 25-JUN-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 7-362553
 ; FILING DATE: 14-SEP-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 16887/845
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136

INFORMATION FOR SEQ ID NO: 66:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 119 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-685-871-66

Query Match 50.6%; Score 42; DB 3; Length 119;
 Best Local Similarity 53.3%; Pred. No. 13;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VQSMREHKAELREK 16
 : : | | | | | | | |
 Db 1 IKEMWARKQELTEK 15

RESULT 3

US-08-685-576-1
 ; Sequence 1, Application US/08685576
 ; Patent No. 5906819
 ; GENERAL INFORMATION:
 ; APPLICANT: Kaibuchi, Koza
 ; APPLICANT: Iwamatsu, Akihiro
 ; APPLICANT: Nakano, Takeshi
 ; APPLICANT: Ito, Masaaki
 ; APPLICANT: Takahashi, No. 5906819uaki
 ; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500

CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/685,576
 ; FILING DATE: 24-JUL-1996
 ; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 7-325129
 ; FILING DATE: 20-NOV-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 8-17150
 ; FILING DATE: 05-JAN-1996
 ; APPLICATION DATA:
 ; APPLICATION NUMBER: JP 8-131206
 ; FILING DATE: 26-APR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bent, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 16887/843

TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1388 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-685-576-1

Query Match 50.6%; Score 42; DB 2; Length 1388;
 Best Local Similarity 53.3%; Pred. No. 1.5e+02;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VQSMREHKAELREK 16
 : : | | | | | | | |
 Db 950 IKEMWARKQELTEK 964

RESULT 4

US-08-685-576-4
 ; Sequence 4, Application US/08685576
 ; Patent No. 5906819
 ; GENERAL INFORMATION:
 ; APPLICANT: Kaibuchi, Koza
 ; APPLICANT: Iwamatsu, Akihiro
 ; APPLICANT: Nakano, Takeshi
 ; APPLICANT: Ito, Masaaki
 ; APPLICANT: Takahashi, No. 5906819uaki
 ; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/685,576

/ FILING DATE: 24-JUL-1996
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 7-325129
/ FILING DATE: 20-NOV-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 8-17150
/ FILING DATE: 05-JAN-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 8-131206
/ FILING DATE: 26-APR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Bent, Stephen A.
/ REGISTRATION NUMBER: 29,768
/ REFERENCE/DOCKET NUMBER: 16887/843
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202)672-5300
/ TELEFAX: (202)672-5399
/ TELEX: 904136
/ INFORMATION FOR SEQ ID NO: 4:
/ LENGTH: 1388 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-685-576-4

Query Match 50.6%; Score 42; DB 2; Length 1388;
Best Local Similarity 53.3%; Pred. No. 1.5e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16
:: |||||
Db 950 IKEMARHQELTEK 964

RESULT 5
US-09-976-594-296
/ Sequence 296, Application US/09976594
/ Patent No. 6673549
/ GENERAL INFORMATION:
/ APPLICANT: Furness, Michael
/ APPLICANT: Buchbinder, Jenny
/ TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
/ FILE REFERENCE: PA-0041 US
/ CURRENT APPLICATION NUMBER: US/09/976,594
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: 60/240,409
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 1143
/ SOFTWARE: PERL Program
/ SEQ ID NO 296
/ LENGTH: 1388
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. 6673549 064987CD1
US-09-976-594-296

Query Match 50.6%; Score 42; DB 4; Length 1388;
Best Local Similarity 53.3%; Pred. No. 1.5e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16
:: |||||
Db 950 IKEMARHQELTEK 964

RESULT 6
US-08-715-204-6
/ Sequence 6, Application US/08715204
/ Patent No. 5874286

/ GENERAL INFORMATION:
/ APPLICANT: Bandman, Olga
/ APPLICANT: Au-Young, Janice
/ APPLICANT: Goli, Surya K.
/ APPLICANT: Hillman, Jennifer.
/ APPLICANT: Zweiger, Gary B.
/ TITLE OF INVENTION: A NOVEL TUMOR PROTEIN
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.
/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: U.S.
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq Version 1.5
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/715,204
/ FILING DATE: Filed Herewith
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J.
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0126 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-855-0555
/ TELEFAX: 415-845-4166
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 257 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 1072344
US-08-715-204-6

Query Match 49.4%; Score 41; DB 2; Length 257;
Best Local Similarity 53.3%; Pred. No. 40;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16
|||::|||
Db 81 VLSARQKHAELRK 95

RESULT 7
US-09-162-597-6
/ Sequence 6, Application US/09162597
/ Patent No. 6043343
/ GENERAL INFORMATION:
/ APPLICANT: Bandman, Olga
/ APPLICANT: Au-Young, Janice
/ APPLICANT: Goli, Surya K.
/ APPLICANT: Hillman, Jennifer.
/ APPLICANT: Zweiger, Gary B.
/ TITLE OF INVENTION: A NOVEL TUMOR PROTEIN
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.
/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: U.S.
/ ZIP: 94304

; TITLE OF INVENTION: Members of the D52 Gene Family
 ; FILE REFERENCE: 1383.0210002
 ; CURRENT APPLICATION NUMBER: US/09/250.609A
 ; CURRENT FILING DATE: 1999-02-17
 ; NUMBER OF SEQ ID NOS: 108
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 144
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-250-609-2

Query Match 47.0%; Score 39; DB 4; Length 144;
 Best Local Similarity 46.7%; Pred. No. 46;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELEK 16
 | : : : : :
 DB 69 VLSAKERHLVETKQ 83

RESULT 11
 US-09-250-611-2
 ; Sequence 2, Application US/09250611
 ; Patent No. 6528283
 ; GENERAL INFORMATION:
 ; APPLICANT: Byrnes, Paul
 ; APPLICANT: Basset, Paul
 ; TITLE OF INVENTION: Members of the D52 Gene Family
 ; FILE REFERENCE: 1383.0210001
 ; CURRENT APPLICATION NUMBER: US/09/250.611
 ; CURRENT FILING DATE: 1999-02-17
 ; NUMBER OF SEQ ID NOS: 108
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 144
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-250-611-2

Query Match 47.0%; Score 39; DB 4; Length 144;
 Best Local Similarity 46.7%; Pred. No. 46;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELEK 16
 | : : : : :
 DB 69 VLSAKERHLVETKQ 83

RESULT 12
 US-09-489-039A-11417
 ; Sequence 11417, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489.039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 11417
 ; LENGTH: 202
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-11417

Query Match 47.0%; Score 39; DB 4; Length 202;
 Best Local Similarity 57.1%; Pred. No. 65;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 SMRERHKAELEK 17
 | : : : : :
 DB 177 SLQRHEALLRELA 190

RESULT 13
 US-08-715-204-1
 ; Sequence 1, Application US/08715204
 ; Patent No. 5874286
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Au-Young, Janice
 ; APPLICANT: Goli, Surya K.
 ; APPLICANT: Hillman, Jennifer.
 ; APPLICANT: Zweiger, Gary B.
 ; TITLE OF INVENTION: A NOVEL TUMOR PROTEIN
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: U.S.
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 1.5
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/715,204
 ; FILING DATE: Filed Herewith
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0126 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 204 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; IMMEDIATE SOURCE:
 ; LIBRARY:
 ; CLONE: Consensus
 US-08-715-204-1

Query Match 47.0%; Score 39; DB 2; Length 204;
 Best Local Similarity 46.7%; Pred. No. 65;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELEK 16
 | : : : : :
 DB 56 VLSAKERHLVETKQ 70

RESULT 14
 US-08-691-814B-10
 ; Sequence 10, Application US/08691814B
 ; Patent No. 5981218
 ; GENERAL INFORMATION:
 ; APPLICANT: Rio, Marie-Christine
 ; APPLICANT: Tomasetto, Catherine
 ; APPLICANT: Basset, Paul
 ; APPLICANT: Byrne, Jennifer
 ; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful

;; TITLE OF INVENTION: as Leukemia Markers and in Breast Cancer Prognosis

;; NUMBER OF SEQUENCES: 124

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

;; STREET: 1100 New York Ave, NW, Suite 600

;; CITY: Washington

;; STATE: DC

;; COUNTRY: USA

;; ZIP: 20005-3934

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patent In Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/691,814B

;; FILING DATE: 31-JUL-1996

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 60/002,183

;; FILING DATE: 09-AUG-1995

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Steffe, Eric K.

;; REGISTRATION NUMBER: 36,688

;; REFERENCE/DOCKET NUMBER: 1383.0090001

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 202-371-2600

;; TELEFAX: 202-371-2543

;; INFORMATION FOR SEQ ID NO: 10:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 204 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; US-08-691-814B-10

Query Match 47.0%; Score 39; DB 2; Length 204;

Best Local Similarity 46.7%; Pred. No. 65;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16

|||:|||||:|::|

Db 56 VLSAKERHVLVEIKQK 70

RESULT 15

US-09-162-597-1

;; Sequence 1, Application US/09162597

;; Patent No. 6043343

;; GENERAL INFORMATION:

;; APPLICANT: Bandman, Olga

;; APPLICANT: Au-Young, Janice

;; APPLICANT: Goli, Surya K.

;; APPLICANT: Hillman, Jennifer.

;; APPLICANT: Zweiger, Gary B.

;; TITLE OF INVENTION: A NOVEL TUMOR PROTEIN

;; NUMBER OF SEQUENCES: 7

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Incyte Pharmaceuticals, Inc.

;; STREET: 3174 Porter Drive

;; CITY: Palo Alto

;; STATE: CA

;; COUNTRY: U.S.

;; ZIP: 94304

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette

;; COMPUTER: IBM Compatible

;; OPERATING SYSTEM: DOS

;; SOFTWARE: FastSeq Version 1.5

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/162,597

;; FILING DATE:

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/715,204

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Billings, Lucy J.

;; REGISTRATION NUMBER: 36,749

;; REFERENCE/DOCKET NUMBER: PF-0126 US

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 415-855-0555

;; TELEFAX: 415-845-4166

;; INFORMATION FOR SEQ ID NO: 1:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 204 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: peptide

;; IMMEDIATE SOURCE:

;; LIBRARY:

;; CLONE: Consensus

;; US-09-162-597-1

Query Match 47.0%; Score 39; DB 3; Length 204;

Best Local Similarity 46.7%; Pred. No. 65;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16

|||:|||||:|::|

Db 56 VLSAKERHVLVEIKQK 70

Search completed: March 9, 2004, 09:28:03

Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 09:25:45 ; Search time 34 Seconds
(without alignments)
105.577 Million cell updates/sec

Title: US-09-998-491-7
Perfect score: 83
Sequence: 1 MVQSMREKHAELREKA 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 83 | 100.0 | 17 | 10 | US-09-998-491-7 |
| 2 | 42 | 50.6 | 258 | 14 | Sequence 5344, Ap |
| 3 | 42 | 50.6 | 447 | 15 | US-10-106-698-5344 |
| 4 | 42 | 50.6 | 1089 | 15 | US-10-104-047-2123 |
| 5 | 42 | 50.6 | 1379 | 14 | US-10-259-194A-264 |
| 6 | 41 | 49.4 | 466 | 14 | US-10-205-219-5 |
| 7 | 41 | 49.4 | 623 | 15 | US-10-032-585-7449 |
| 8 | 40.5 | 48.8 | 375 | 9 | US-10-104-047-3378 |
| 9 | 40 | 48.2 | 517 | 9 | US-09-815-242-10928 |
| 10 | 40 | 48.2 | 519 | 9 | US-09-815-242-5387 |
| 11 | 40 | 48.2 | 519 | 9 | US-09-815-242-12331 |
| 12 | 40 | 48.2 | 529 | 14 | US-10-104-047-2123 |
| 13 | 40 | 48.2 | 668 | 15 | US-10-369-493-18727 |
| 14 | 40 | 48.2 | 854 | 15 | US-10-428-487-26 |
| 15 | 39 | 47.0 | 144 | 9 | US-09-250-611-2 |

| | | | | | |
|----|------|------|------|----|---------------------|
| 16 | 39 | 47.0 | 204 | 9 | US-09-250-611-4 |
| 17 | 39 | 47.0 | 204 | 9 | US-09-250-611-9 |
| 18 | 39 | 47.0 | 336 | 15 | US-10-369-493-18231 |
| 19 | 39 | 47.0 | 853 | 14 | US-10-156-761-8930 |
| 20 | 39 | 47.0 | 1109 | 9 | US-09-529-063-55 |
| 21 | 39 | 47.0 | 1109 | 14 | US-10-414-378-55 |
| 22 | 38.5 | 46.4 | 168 | 9 | US-09-791-118A-6 |
| 23 | 38.5 | 46.4 | 168 | 14 | US-10-232-188-3 |
| 24 | 38.5 | 46.4 | 169 | 15 | US-10-133-628-13 |
| 25 | 38 | 45.8 | 63 | 14 | US-10-029-386-28566 |
| 26 | 38 | 45.8 | 151 | 15 | US-10-369-493-983 |
| 27 | 38 | 45.8 | 307 | 15 | US-10-264-237-2562 |
| 28 | 38 | 45.8 | 347 | 15 | US-10-094-749-2106 |
| 29 | 38 | 45.8 | 993 | 15 | US-10-443-735-4 |
| 30 | 38 | 45.8 | 1049 | 14 | US-10-317-835-16 |
| 31 | 38 | 45.8 | 1145 | 15 | US-10-369-493-20484 |
| 32 | 38 | 45.8 | 1152 | 15 | US-10-369-493-12292 |
| 33 | 38 | 45.8 | 1173 | 15 | US-10-369-493-5025 |
| 34 | 38 | 45.8 | 1179 | 9 | US-09-815-242-13262 |
| 35 | 38 | 45.8 | 1179 | 9 | US-09-815-242-13608 |
| 36 | 38 | 45.8 | 2099 | 14 | US-10-128-714-3290 |
| 37 | 38 | 45.8 | 2405 | 14 | US-10-128-714-8290 |
| 38 | 37.5 | 45.2 | 829 | 9 | US-09-815-242-11216 |
| 39 | 37 | 44.6 | 78 | 14 | US-10-029-386-29141 |
| 40 | 37 | 44.6 | 185 | 9 | US-09-250-611-11 |
| 41 | 37 | 44.6 | 308 | 14 | US-10-043-487-393 |
| 42 | 37 | 44.6 | 324 | 15 | US-10-369-493-10612 |
| 43 | 37 | 44.6 | 373 | 15 | US-10-120-835-42 |
| 44 | 37 | 44.6 | 427 | 15 | US-10-443-201-8 |
| 45 | 37 | 44.6 | 430 | 10 | US-09-882-227-516 |

ALIGNMENTS

RESULT 1

US-09-998-491-7
; Sequence 7, Application US/09998491
; Publication No. US2003016529A1
; GENERAL INFORMATION:
; APPLICANT: Mileusnic, Radmilla
; TITLE OF INVENTION: Polypeptides and their Uses
; FILE REFERENCE: 3578-120
; CURRENT APPLICATION NUMBER: US/09/998,491
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: GB 0109558.7
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: GB 0120084
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 17
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 16-mer polypeptide
US-09-998-491-7

Query Match 100.0%; Score 83; DB 10; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQSMREKHAELREKA 17

Db 1 MVQSMREKHAELREKA 17

RESULT 2

US-10-106-698-5344
; Sequence 5344, Application US/10106698
; Publication No. US20030109690A1

| | | | | |
|--|---------|-----------------|-----|------|
| | Matches | 8; Conservative | 2; | Mis- |
| Qy | 2 | VQSMREHKAEIREK | 16 | |
| | : | : : | | |
| Dd | 941 | IKEMARHQSLTEK | 955 | |
| RESULT 6 | | | | |
| US-10--032-585-7449 | | | | |
| ; Sequence 7449, Application US/10032585 | | | | |
| ; Publication No. US2003018095A1 | | | | |
| ; GENERAL INFORMATION: | | | | |

; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7449
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7449

Query Match 49.4%; Score 41; DB 14; Length 466;
Best Local Similarity 72.7%; Pred. No. 2.3e+02;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 RERHKAELREK 16
|:|:|:|:|:|:|:
Db 151 ROREKAELRK 161

RESULT 7
US-10-104-047-3378
; Sequence 3378, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3378
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3378

Query Match 49.4%; Score 41; DB 15; Length 623;
Best Local Similarity 53.3%; Pred. No. 3.1e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 QSMRERHKAELREKA 17
|:|:|:|:|:|:|:
Db 464 QGLRERQRRELEKA 478

RESULT 8
US-09-815-242-10928
; Sequence 10928, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10928
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10928

Query Match 48.8%; Score 40.5; DB 9; Length 375;
Best Local Similarity 73.3%; Pred. No. 2.2e+02;
Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 MVOSMEER-HKAELR 14
|:|:|:|:|:|:|:
Db 128 MVQWMEERAEGLR 142

RESULT 9
US-09-815-242-5387
; Sequence 5387, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5387
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5387

Query Match 48.2%; Score 40; DB 9; Length 517;
Best Local Similarity 57.1%; Pred. No. 3.7e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 QSMRHHKAELEK 16
| :||: :||||:
DB 64 QILREQTEALELR 77

RESULT 10

US-09-815-242-12331
; Sequence 12331, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12331
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12331

Query Match 48.2%; Score 40; DB 9; Length 519;
Best Local Similarity 57.1%; Pred. No. 3.7e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 QSMRHHKAELEK 16
| :||: :||||:
DB 66 QILREQTEALELR 79

RESULT 11

US-09-815-242-12869
; Sequence 12869, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12869
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12869

Query Match 48.2%; Score 40; DB 9; Length 519;
Best Local Similarity 57.1%; Pred. No. 3.7e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 QSMRHHKAELEK 16
| :||: :||||:
DB 66 QILREQTEALELR 79

RESULT 12

US-10-301-997-74
; Sequence 74, Application US/10301997
; Publication No. US20030148346A1
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/301,997
; FILING DATE: 22-NO. US20030148346A1-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/527,431
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US/08/887,534
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-301-997-74

```

Query Match 48.2%; Score 40; DB 14; Length 529;
Best Local Similarity 57.1%; Pred. No. 3.8e+02;
Matches 8; Conservative 4; Mismatches 2; Indels

Qy 3 QSMRERHKAELREK 16
| : | : : | | | :
Db 76 QILREOTEAE LRE R 89

```

RESULT 13
US-10-369-493-18727
/ Sequence 18727, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-101520521B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 18727
/ LENGTH: 668
/ TYPE: PRT
/ ORGANISM: Anabaena PCC7120
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(668)
/ OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-18727

```

Query Match 48.2%; Score 40; DB 15; Length 668;
Best Local Similarity 58.3%; Pred. NO. 4.8e+02;
Matches 7; Conservative 3; Mismatches 2; Indels

Qy 3 QSMRERHKAELR 14
| : | | : | | :
Db 222 QELAERHQAELK 233

```

RESULT 14
US-10-428-487-26
; Sequence 26, Application US/10428487
; Publication No. US20040006780A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA K.
; APPLICANT: GERBER, HANS-PETER
; TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM
; FILE REFERENCE: 09800080-0103
; CURRENT APPLICATION NUMBER: US/10/428,487
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: 09/815,153
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,201
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 854
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-428-487-26

```

Query Match 48.2%; Score 40; DB 15; Length 854;
Best Local Similarity 50.0%; Pred. No. 6.2e+02;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 VQSMRERHKAELRE 15
| | : : | | : : | :
Db 240 VCSVKERHASEMD 253

```

RESULT 15
US-09-250-611-2
; Sequence 2, Application US/09250611
; Patent No. US20020143161a1
; GENERAL INFORMATION:
; APPLICANT: Byrnie, Jennifer A.
; APPLICANT: Basset, Paul
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210001
; CURRENT APPLICATION NUMBER: US/09/250.611
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 144
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-250-611-2

```

```
Query Match      47.0%; Score 39; DB 9; Length 144;
Best Local Similarity 46.7%; Pred. No. 1.4e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
```

Qy 2 VQSMRERHKAELREK 16
Db 69 VLSAKERHLEIKOK 83

Search completed; March 9, 2004, 09:28:50
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 09:24:30 ; Search time 20 Seconds
(without alignments)
81.763 Million cell up

Title: US-09-998-491-7
Perfect score: 83
Sequence: 1 MVOSMRERHKAELREKA 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : PIR_78:**
1:  _pir1:**
2:  _pir2:**
3:  _pir3:**
4:  _pir4:**
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | |
|------------|-------|-------------|--------|----|--------|--------------------|--|
| | | | | | | | |
| 1 | 43.5 | 52.4 | 279 | 2 | T19828 | hypothetical prote | |
| 2 | 43 | 51.8 | 309 | 2 | T21195 | hypothetical prote | |
| 3 | 43 | 51.8 | 987 | 2 | T10562 | hypothetical prote | |
| 4 | 42 | 50.6 | 433 | 2 | T36122 | probable integrase | |
| 5 | 42 | 50.6 | 1029 | 2 | T20576 | hypothetical prote | |
| 6 | 42 | 50.6 | 1388 | 2 | S70633 | serine/threonine-s | |
| 7 | 42 | 50.6 | 1388 | 2 | S74245 | serine/threonine-s | |
| 8 | 41 | 49.4 | 257 | 2 | E89632 | protein F3E6.1 [i | |
| 9 | 41 | 49.4 | 284 | 1 | LNRTL | hepatic lectin - r | |
| 10 | 41 | 49.4 | 553 | 1 | DJBPD1 | DNA-directed DNA p | |
| 11 | 41 | 49.4 | 3225 | 2 | I52300 | giantin - human | |
| 12 | 41 | 49.4 | 3259 | 1 | A56539 | giantin - human | |
| 13 | 40.5 | 48.8 | 552 | 2 | AF1301 | hypothetical prote | |
| 14 | 40.5 | 48.8 | 552 | 2 | AF1673 | hypothetical prote | |
| 15 | 40 | 48.2 | 97 | 2 | S47150 | NAH2 dehydrogenas | |
| 16 | 40 | 48.2 | 124 | 2 | T21183 | hypothetical prote | |
| 17 | 40 | 48.2 | 204 | 2 | A83440 | hypothetical prote | |
| 18 | 40 | 48.2 | 288 | 2 | H90392 | hypothetical prote | |
| 19 | 40 | 48.2 | 387 | 2 | S02708 | conserved hypothet | |
| 20 | 40 | 48.2 | 396 | 2 | S13251 | troponin T - fruit | |
| 21 | 40 | 48.2 | 477 | 2 | T37791 | troponin T - fruit | |
| 22 | 40 | 48.2 | 519 | 2 | A89903 | probable rna polym | |
| 23 | 40 | 48.2 | 530 | 2 | T52508 | conserved hypothet | |
| 24 | 40 | 48.2 | 658 | 2 | F84254 | hypothetical prote | |
| 25 | 40 | 48.2 | 918 | 2 | C96829 | hypothetical prote | |
| 26 | 40 | 48.2 | 1132 | 2 | T00259 | unknown protein F1 | |
| 27 | 40 | 48.2 | 1286 | 2 | A12195 | hypothetical prote | |
| 28 | 40 | 48.2 | 1976 | 2 | A59522 | two-component hybr | |
| 29 | 40 | 48.2 | 2168 | 2 | T30171 | myosin heavy chain | |
| | | | | | | nipin - mouse | |

ALIGNMENTS

RESULT 1

T19828
 Hypothetical protein C38D4.7 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T19828
 R:Colles, I.
 Submitted to the EMBL Data Library, October 1994
 A:Reference number: Z19183
 A:Accession: T19828
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-279 <WIL>
 A:Cross-references: EMBL:Z46241; PTDN:CAA6321.i; GSPDB:GN00021; CBSP:C38D4.7
 A:Experimental source: clone C38D4
 C:Genetics:
 A:Gene: CESP:C38D4.7
 A:Map position: 3
 A:Introns: 14/3; 76/3; 213/2; 231/3

Query Match 52.4%; Score 43.5; DB 2; Length 279;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 10; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Qy 3 QSMRERHKAELREKA 17
: : : : :
Db 110 RSIRREEHRAE-REKA 123

RESULT 2

ZRS001.2
 Z21195
 A:Hyphothetical protein F21D5.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 S:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: Z21195
 R:Berks, M.
 Submitted to the EMBL Data Library, September 1995
 A:Reference number: Z19389
 A:Accession: Z21195
 A:Status: preliminary; translated from GE/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-309 <Nil>
 A:Cross-references: EMBL:Z54271; PIDN:CAA91033.1; GSPDB:GN00022; CESP:F21D5.2
 A:Experimental source: Clone F21D5
 C:Genetics:
 A:Gene: CESP:F21D5.2
 A:Map position: 4
 A:Introns: 26/1; 49/3; 160/3; 265/3

```
Query Match      51.8%; Score 43; DB 2; Length 309;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

QY 1 VQSMRERHKAELREK 15
 Db 53 MEKEDRRHKAELREK 67

RESULT 3
 T10562
 hypothetical protein F25E4.40 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
 C:Accession: T10562
 R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro
 submitted to the Protein Sequence Database, June 1999
 A:Reference number: Z16533
 A:Accession: T10562
 A:Molecule type: DNA
 A:Residues: 1-987 <BEV>
 A:Cross-references: EMBL:AL050399; GSPDB:GN00062; ATSP:F25E4.40
 A:Experimental source: cultivar Columbia; BAC clone F25E4
 C:Genetics:
 A:Gene: ATSP:F25E4.40
 A:Map position: 4
 A:Introns: 16/1; 176/3; 248/3; 371/3; 381/3; 457/3; 511/3; 585/3; 654/3; 723/3; 750/3; 8

Query Match 51.8%; Score 43; DB 2; Length 987;
 Best Local Similarity 46.7%; Pred. No. 87;
 Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16
 Db 726 VELSKERHSDLKEK 740

RESULT 4
 T36122
 probable integral membrane protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T36122
 R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, April 1999
 A:Reference number: Z21597
 A:Accession: T36122
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-433 <MUR>
 A:Cross-references: EMBL:AL049707; PIDN:CAB41288.1; GSPDB:GN00070; SCOEDB:SCB15.19
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCB15.19

Query Match 50.6%; Score 42; DB 2; Length 433;
 Best Local Similarity 50.0%; Pred. No. 56;
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VQSMRERHKAELREK 16
 Db 415 MVKAFSEHRELQEE 430

RESULT 5
 T02576
 hypothetical protein At2g39260 [imported] - Arabidopsis thaliana
 N:Alternate names: hypothetical protein T16B24.10
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C:Accession: T02576; B84815
 R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
 submitted to the EMBL Data Library, August 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
 A:Reference number: Z14679
 A:Accession: T02576

A:Status: translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-1029 <ROU>
 A:Cross-references: EMBL:AC004697; NID:g3402671; PID:g3402679
 A:Experimental source: cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shet, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: B84815
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1029 <STO>
 A:Cross-references: GB:AE002093; NID:g3402679; PIDN:AAC28982.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g39260; T16B24.10
 A:Map position: 2
 A:Introns: 72/3; 104/3; 147/2; 205/2; 232/3; 330/3; 396/3; 430/3; 514/3; 587/3; 644/3; 70

Query Match 50.6%; Score 42; DB 2; Length 1029;
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 VQSMRERHKAELREK 16
 Db 946 LVQSTQKRAELEK 961

RESULT 6
 S70633
 serine/threonine-specific protein kinase (EC 2.7.1.1.-), Rho-associated - bovine
 N:Alternate names: Rho-associated protein kinase
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 24-Sep-1999
 C:Accession: S70633; S77694
 R:Matsum, T.; Anano, M.; Yamamoto, T.; Chihara, K.; Nakafuku, M.; Ito, M.; Nakano, T.; O
 EMBO J. 15, 2208-2216, 1996
 A>Title: Rho-associated kinase, a novel serine/threonine kinase, as a putative target for
 A:Reference number: S70633; MUID:96208507; PMID:8641286
 A:Accession: S70633
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1388 <MAT>
 A:Cross-references: EMBL:U036909; NID:gl326077; PIDN:AAC48567.1; PID:gl326078
 A:Accession: S77694
 A:Molecule type: protein
 A:Residues: 1-18;30-34;36-44;58-64;133-140;248-252;291-295;327-347;350-360;366-37
 -1070 <MAT2>
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase C zinc-
 C:Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase
 F:90-354/Domain: protein kinase homology <KIN>
 F:98-106/Region: protein kinase ATP-binding motif
 F:1261-1315/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 50.6%; Score 42; DB 2; Length 1388;
 Best Local Similarity 53.3%; Pred. No. 1.7e+02;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16
 Db 950 IKEMWARKQLTEK 964

RESULT 7
 S74245
 serine/threonine-specific protein kinase (EC 2.7.1.1.-) isoform II, Rho-associated - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 24-Sep-1999
 C:Accession: S74245
 R:Nakagawa, O.; Fujisawa, K.; Ishizaki, T.; Saito, Y.; Nakao, K.; Narumiya, S.
 FEBS Lett. 392, 189-193, 1996

A;Title: ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein
A;Reference number: S74244; MUID:96368048; PMID:8772201
A;Accession: S74245
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1388 <NAK>
A;Cross-references: EMBL:U58513; NID:q1514697; PIDN:RAC53133.1; PID:q1514698
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase C zinc
C;Keywords: phosphotransferase; serine/threonine-specific protein kinase
F;90-354/Domain: protein kinase homology <KIN>
F;1261-1315/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 50.6%; Score 42; DB 2; Length 1388;
Best Local Similarity 53.3%; Pred. No. 1.7e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16
: : | | | | | | | |
Db 950 IKEMARHKEQTEK 964

RESULT 8
E89632
A;Title: F13B6.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
C;Accession: E89632
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E89632
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-257 <STO>
A;Cross-references: GB:chr_X; PIDN:CAA92122.1; PID:g3875842; GSPDB:GN00028; CESP:F13B6.1
A;Gene: F13B6.1
A;Map position: X

Query Match 49.4%; Score 41; DB 2; Length 257;
Best Local Similarity 53.3%; Pred. No. 48;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16
: : | | | | | | | |
Db 81 VLSARQKHAELKPK 95

RESULT 9
LNRTL
A;Title: Rat liver asialoglycoprotein receptor
A;Accession: A92497
A;Molecule type: DNA
A;Residues: 1-284 <LEU>
A;Cross-references: GB:K02817; NID:g206646; PIDN:AAA42037.1; PID:g206647
R;Holland, E.C.; Leung, J.O.; Drickamer, K.
Proc. Natl. Acad. Sci. U.S.A. 81, 7338-7342, 1984
A;Title: Rat liver asialoglycoprotein receptor lacks a cleavable NH-2-terminal signal se
A;Reference number: A94020; MUID:85063786; PMID:6095287
A;Accession: A94020
A;Molecule type: mRNA
A;Residues: 1-60, 'R', 62-210 <HOL>

A;Experimental source: clone 22; clone 1
A;Accession: B94020
A;Molecule type: mRNA
A;Residues: 92-284 <HO2>
A;Note: clone 22 codes for a terminator at residue 210
R;Watts, C.
BioSci. Rep. 6, 527-534, 1986
A;Title: Isolation and expression of cDNA clones for a rat liver asialoglycoprotein recep
A;Reference number: A54727; MUID:87026895; PMID:2945599
A;Accession: A54727
A;Molecule type: mRNA
A;Residues: 12-284 <WAT>
A;Cross-references: GB:M21770; NID:g202985; PIDN:AAA40764.1; PID:g202988
A;Experimental source: liver
C;Comment: Two types of rat hepatic lectin have been identified, RHL-1 and RHL-2/3, havir
C;Comment: After removal of sialic acid monomers from the complex carbohydrate moieties
C;Comment: The unusual orientation of this protein across the membrane is postulated to
C;Genetics:
A;Introns: 23/1; 62/1; 94/1; 118/1; 147/1; 197/3; 233/2
C;Superfamily: hepatic lectin; C-type lectin homology
C;Keywords: endocytosis; glycoprotein; lectin; receptor; transmembrane protein
F;2-284/Product: hepatic lectin #status predicted <WAT>
F;2-39/Domain: intracellular #status predicted <INT>
F;40-60/Domain: transmembrane #status predicted <TM>
F;61-284/Domain: extracellular #status predicted <EXT>
F;153-276/Domain: C-type lectin homology <UCH>
F;75,78,146/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 49.4%; Score 41; DB 1; Length 284;
Best Local Similarity 46.7%; Pred. No. 53;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 MVQSMRERHKAELRE 15
: : | | | | | | | |
Db 103 LVESQLEKHQEDLRE 117

RESULT 10
DUBPD1
A;Title: DNA-directed DNA polymerase (EC 2.7.7.7) - phage PRD1
C;Species: phage PRD1
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 11-Jun-1999
C;Accession: B27328; B40477; A39970
R;Savilahti, H.; Bamford, D.H.
Gene 57, 121-130, 1987
A;Title: The complete nucleotide sequence of the left very early region of Escherichia c
A;Reference number: A27328; MUID:88112855; PMID:3322943
A;Accession: B27328
A;Molecule type: DNA
A;Residues: 1-553 <SAV>
A;Cross-references: GB:M22161; NID:g215750; PIDN:AAA32450.1; PID:g215752
R;Bamford, J.K.H.; Haenninen, A.L.; Pakula, T.M.; Ojala, P.M.; Kalkkinen, N.; Frilander,
Virology 183, 658-676, 1991
A;Title: Genome organization of membrane-containing bacteriophage PRD1.
A;Reference number: A40477; MUID:91306449; PMID:1853567
A;Accession: B40477
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-553 <BAM>
A;Cross-references: GB:M69077; NID:g215765; PIDN:AAA32456.1; PID:g215767
R;Jung, G.; Leavitt, M.C.; Heien, J.C.; Ito, J.
Proc. Natl. Acad. Sci. U.S.A. 84, 8287-8291, 1987
A;Title: Bacteriophage PRD1 DNA polymerase: evolution of DNA polymerases.
A;Reference number: A39970; MUID:88068579; PMID:3479792
A;Accession: A39970
A;Molecule type: DNA
A;Residues: 1-553 <JUN>
A;Cross-references: NID:g215760; PIDN:AAA32452.1; PID:g215761
C;Comment: This DNA polymerase primes itself by forming a phosphodiester bond between 5'-
C;Superfamily: phage PRD1 DNA-directed DNA polymerase
C;Keywords: DNA binding; nucleotidyltransferase

```

Query Match      49.4%; Score 41; DB 1; Length 553;
Best Local Similarity 53.3%; Pred. No. 1e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1  MVQSMREHKAELREK 15
DB      130  MERETREQHKAELILE 144

RESULT 11
I52300
gi|152300|ref|NM_001126801.1|g|152300|c|152300|
gi|152300|ref|NM_001126801.1|g|152300|c|152300|
N:Alternate names: gop372
C:Species: Homo sapiens (man)
C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C:Accession: I52300
R:Sohda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A>Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A:Reference number: I52300; MUID:95100974; PMID:7802676
A:Accession: I52300
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: GB:D25542; NID:9662389; PIDN:BA05025.1; PID:9808869
C:Superfamily: giantin

Query Match      49.4%; Score 41; DB 2; Length 3225;
Best Local Similarity 46.2%; Pred. No. 5.6e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      4  SMRERHKAELREK 16
DB      331  ALQKHKAEMEEK 343

RESULT 12
A56539
gi|56539|ref|NM_001126801.1|g|56539|c|56539|
N:Alternate names: macrogolglin
C:Species: Homo sapiens (man)
C>Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
C:Accession: A56539; S37536
R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A>Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein
A:Reference number: A56539; MUID:94187728; PMID:7511208
A:Accession: A56539
A:Molecule type: mRNA
A:Residues: 1-3259 <SEE>
A:Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715
C:Genetics:
A:Gene: GDB:GOLG1; GCP: GCP371
A:Cross-references: GDB:454958
A:Map position: 3q13.31-3q13.31
C:Superfamily: giantin
C:Keywords: coiled coil; Golgi apparatus; transmembrane protein
F/3238-3254/Domain: transmembrane #status predicted <TM>

Query Match      49.4%; Score 41; DB 1; Length 3259;
Best Local Similarity 46.2%; Pred. No. 5.6e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      4  SMRERHKAELREK 16
DB      365  ALQKHKAEMEEK 377

RESULT 13
AF1301
gi|1301|ref|AF1301|g|1301|c|1301|
N:hypothetical protein lmo1814 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

```

```

C:Accession: AF1301
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A>Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1301
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-552 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99892.1; PID:g16411268; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo1814
C:Superfamily: Mycoplasma genitalium hypothetical protein MG369

Query Match      48.8%; Score 40.5; DB 2; Length 552;
Best Local Similarity 43.5%; Pred. No. 1.2e+02;
Matches 10; Conservative 4; Mismatches 2; Indels 7; Gaps 1;

QY      2  VQSMREH-----KAELEKA 17
DB      312  VENNREQHNEIVGDDKAPAKEKA 334

RESULT 14
AF1673
gi|1673|ref|AF1673|g|1673|c|1673|
N:hypothetical protein homolog lln1928 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1673
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A>Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1673
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-552 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97158.1; PID:g16414429; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lln1928
C:Superfamily: Mycoplasma genitalium hypothetical protein MG369

Query Match      48.8%; Score 40.5; DB 2; Length 552;
Best Local Similarity 43.5%; Pred. No. 1.2e+02;
Matches 10; Conservative 4; Mismatches 2; Indels 7; Gaps 1;

QY      2  VQSMREH-----KAELEKA 17
DB      312  VENNREQHNEIVGDDKAPAKEKA 334

RESULT 15
S47150
gi|47150|ref|S47150|g|47150|c|47150|
N:NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) 14K chain - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 23-Nov-1994 #sequence_revision 13-Mar-1997 #text_change 03-Jun-2002
C:Accession: S47150
R:Nehls, U.; Kruehl, C.; Weiss, H.
submitted to the EMBL Data Library, December 1992
A:Reference number: S47150
A:Accession: S47150
A>Status: preliminary
A:Molecule type: DNA

```

A;Residues: 1-97 <NEH>
A;Cross-references: EMBL:Z18945; NID:g499314; PID:g499315
C;Keywords: NAD; oxidoreductase

Query Match 48.2%; Score 40; DB 2; Length 97;
Best Local Similarity 53.3%; Fred. No. 27;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 VQSMRERHKALREK 16
|| : ||| : || :
Db 55 VQGVDERHEALQER 69

Search completed: March 9, 2004, 09:27:28
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: March 9, 2004, 09:24:29 ; Search time 11 Seconds
(without alignments)
80.472 Million cell updates/sec

Title: US-09-998-491-7

Perfect score: 83

Sequence: 1 MVQSMRRHRKAELEKA 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------|---------------------|
| 1 | 43 | 51.8 | 536 | 1 SNW1_HUMAN | Q13573 homo sapien |
| 2 | 43 | 51.8 | 987 | 1 IF3A_ARATH | Q13573 arabidopsis |
| 3 | 42.5 | 51.2 | 329 | 1 IMP3_HUMAN | Q9NUV9 homo sapien |
| 4 | 42 | 50.6 | 1583 | 1 CC2_HUMAN | Q8IWI2 homo sapien |
| 5 | 42 | 50.6 | 5560 | 1 SPEN_DRONE | Q8SX83 drosophila |
| 6 | 41 | 49.4 | 195 | 1 YZG1_CABEL | P55326 caenorhabdi |
| 7 | 41 | 49.4 | 283 | 1 LECH_RAT | P02706 rattus norv |
| 8 | 41 | 49.4 | 553 | 1 DPOL_EPRD | P10479 bacterioph |
| 9 | 41 | 49.4 | 3259 | 1 G0B1_HUMAN | Q14789 homo sapien |
| 10 | 40 | 48.2 | 396 | 1 TRT_DROME | P19351 drosophila |
| 11 | 40 | 48.2 | 477 | 1 TP81_SCHPO | Q13745 schizosacch |
| 12 | 40 | 48.2 | 519 | 1 YC86_STAMM | Q99UI7 staphylococ |
| 13 | 40 | 48.2 | 854 | 1 VP41_HUMAN | P49754 homo sapien |
| 14 | 40 | 48.2 | 1976 | 1 MYHA_BOVIN | Q27991 bos taurus |
| 15 | 40 | 48.2 | 1976 | 1 MYHA_RAT | Q35580 rattus norv |
| 16 | 40 | 48.2 | 1976 | 1 MYHA_RAT | Q9J1L0 rattus norv |
| 17 | 40 | 48.2 | 3305 | 1 APPL_MANSE | Q25490 manduca sex |
| 18 | 39.5 | 47.6 | 2238 | 1 GOA4_MOUSE | Q91VW5 mus musculu |
| 19 | 39 | 47.0 | 156 | 1 R57_BRUME | Q9YHP4 bruceella me |
| 20 | 39 | 47.0 | 156 | 1 R57_BRUME | Q9YHP4 bruceella me |
| 21 | 39 | 47.0 | 204 | 1 TD53_HUMAN | Q8G074 bruceella su |
| 22 | 39 | 47.0 | 204 | 1 TD53_HUMAN | Q8G074 bruceella su |
| 23 | 39 | 47.0 | 242 | 1 YB38_PSEAE | Q54818 mus musculu |
| 24 | 39 | 47.0 | 263 | 1 RM41_YEAST | Q9HVI1 pseudomonas |
| 25 | 39 | 47.0 | 305 | 1 GAAB_HAHLN1 | P32387 saccharomyc |
| 26 | 39 | 47.0 | 336 | 1 FEN_THEAC | Q9HP33 halobacteri |
| 27 | 39 | 47.0 | 365 | 1 D41_METMA | Q9HJ04 thermoplasma |
| 28 | 39 | 47.0 | 549 | 1 ES11_RAT | Q9Y959 rattus norv |
| 29 | 39 | 47.0 | 643 | 1 YK03_CABEL | P10959 methanosarc |
| 30 | 39 | 47.0 | 726 | 1 HS9A_BRARE | P34304 caenorhabdi |
| 31 | 38 | 45.8 | 98 | 1 Y126_METUA | Q90474 brachydanio |
| 32 | 38 | 45.8 | 151 | 1 NDK_ARCFU | Q57590 methanococc |
| 33 | 38 | 45.8 | 184 | 1 TD52_RABIT | Q29491 archaeoglob |
| | | | | | Q95212 cryotolagob |

| RESULT 1 | SNW1_HUMAN | STANDARD; | PRT; | 536 AA. |
|----------|--|-----------|------|---------|
| ID | SNW1_HUMAN | STANDARD; | PRT; | 536 AA. |
| AC | Q13573; Q13483; | | | |
| DT | 01-NOV-1997 (Rel. 35, Created) | | | |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) | | | |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) | | | |
| DE | Nuclear protein Skip (Ski-interacting protein) (SNW1 protein) (Nuclear | | | |
| DE | receptor coactivator NCoA-62). | | | |
| GN | SNW1 OR SKIP. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| EX | MEDLINE=98227980; PubMed=9569025; | | | |
| RA | Dahl R., Wani B., Hayman M.J.; | | | |
| RT | "The Ski oncoprotein interacts with Skip, the human homolog of | | | |
| RT | Drosophila BX42."; | | | |
| RL | Oncogene 16:1579-1586(1998). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| EX | MEDLINE=98298162; PubMed=9632709; | | | |
| RA | Baudino T.A., Kraichely D.M., Jefcoat S.C. Jr., Winchester S.K., | | | |
| RA | Partridge N.C., Macdonald P.N.; | | | |
| RT | "Isolation and characterization of a novel coactivator protein, | | | |
| RT | NCoA-62, involved in vitamin D-mediated transcription."; | | | |
| RL | J. Biol. Chem. 273:16434-16441(1998). | | | |
| RN | [3] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Rowen L., Madan A., Qin S., Abbasi N., Baradaran L., Birditt B., | | | |
| RA | Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R., | | | |
| RA | Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.; | | | |
| RT | "Sequencing of human chromosome 14."; | | | |
| RL | Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases. | | | |
| RN | [4] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | TISSUE=Cervix; | | | |
| EX | MEDLINE=92388257; PubMed=12477932; | | | |
| RA | Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., | | | |
| RA | Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., | | | |
| RA | Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., | | | |
| RA | Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., | | | |
| RA | Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., | | | |
| RA | Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., | | | |
| RA | Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., | | | |
| RA | Raha S.S., McQuillan N.A., Peters G.J., Abramson R.D., Mullany S.J., | | | |
| RA | Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H., | | | |
| RA | Richards S., Worley K.C., Halle S., Garcia A.M., Gay L.J., Hulyk S.W., | | | |
| RA | Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., | | | |
| RA | Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., | | | |
| RA | Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., | | | |
| RA | Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., | | | |
| RA | Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., | | | |
| RA | Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., | | | |

ALIGNMENTS

O80840 arabidopsis
O29986 archaeoglob
Q96f15 homo sapien
P93398 nicotiana t
P48386 camellia si
P45970 caenorhabdi
P06198 schistosoma
Q04500 saccharomyc
P44323 haemophilus
Q8y1y5 ralsstonia s
Q62393 mus musculu
Q9y8t2 aeropyrum p

34 38 45.8 246 1 PMN_ARATH
35 38 45.8 303 1 GAAB_ARCFU
36 38 45.8 307 1 IMP3_HUMAN
37 38 45.8 377 1 GBB2_TOBAC
38 38 45.8 389 1 CHSI_CAMSI
39 38 45.8 821 1 LINS_CABEL
40 38 45.8 866 1 MYSP_SCHMA
41 38 45.8 899 1 YMJ3_YEAST
42 37.5 45.2 829 1 LF2_HAEN
43 37 44.6 159 1 RECY_RALSO
44 37 44.6 185 1 TD52_MOUSE
45 37 44.6 345 1 TRPD_AERPE

DR SMART; SM00088; PINT; 1.
 KW Initiation factor; Protein biosynthesis.
 FT DOMAIN 436 510
 SQ SEQUENCE 987 AA; 114298 MW; F38BA715209D55F8 CRC64;

Query Match 51.8%; Score 43; DB 1; Length 987;
 Best Local Similarity 46.7%; Pred. No. 46;
 Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELREK 16
 DB 726 VELSKERHESDLKEK 740

RESULT 3

IMP4_HUMAN IMP4_HUMAN STANDARD; PRT; 329 AA.
 AC Q9NUV9;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DE Immunity-associated protein 4 (Immunity-associated nucleotide 1 protein) (hIAN1) (MSTP062).
 GN IMAF4 OR IAN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=T-cell lymphoma;
 RX MEDLINE=21961548; PubMed=11964296;
 RA Canbot M., Aresta S., Kahn-Perles B., de Gunzburg J., Romeo P.-H.;
 RT "Human immune associated nucleotide 1: a member of a new guanidine triphosphatase family expressed in resting T and B cells.";
 RL Blood 99:3293-3301(2002).
 [2]
 SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RA Liu Y.Q., Liu B., Zhao B., Wang X.Y., Song L., Ye J., Sheng H.,
 RA Gao Y., Zhang C.L., Zhang J., Wei Y.J., Sun Y.H., Jiang Y.X.,
 RA Gao X.W., Liu S., Liu L.S., Ding J.F., Gao R.L., Wu Q.Y.,
 RA Qulang B.Q., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 [3]
 SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
 RA Nakamura Y., Nagahari K., Magubo Y., Sasaki N.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F.F., Bhat N.K.,
 RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Heish N.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Heaton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitting R., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grinstead J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -| FUNCTION: Exhibits intrinsic GTPase activity. Shows a higher affinity for GDP over GTP (about 12-fold higher), and binding shows an absolute requirement for magnesium.
 CC -| TISSUE SPECIFICITY: Highly expressed in spleen and peripheral blood leukocytes that contain mostly T and B lymphocytes.
 CC Expressed specifically in resting T and B lymphocytes and expression significantly decreases during B- or T-lymphocyte activation. Expressed at lower levels in thymus, ovary, colon and small intestine.
 CC -| SIMILARITY: Belongs to the GTP-binding IAN family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF117333; AAC15308.1; -;
 DR EMBL; AK001972; BAA92010.1; -;
 DR EMBL; BC020657; AAH20657.1; -;
 DR MIM; 608087; -;
 DR InterPro; IPR006703; AIG1.
 DR PFam; PF04548; AIG1; 1.
 DR GTP-binding; Coiled coil; Polymorphism.
 KW NP BIND 37 44 GTP (POTENTIAL).
 FT NP BIND 85 88 GTP (POTENTIAL).
 FT NP BIND 156 158 GTP (POTENTIAL).
 FT DOMAIN 188 300 COILED COIL (POTENTIAL).
 FT VARIANT 128 128 E -> D (in dbSNP:22931172).
 FT /FTID=VAR 017306.
 SQ SEQUENCE 329 AA; 37534 MW; 9D64BA4FB1C5DF72 CRC64;

Query Match 51.2%; Score 42.5; DB 1; Length 329;
 Best Local Similarity 62.5%; Pred. No. 19;
 Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 3 QSMRERHKAEL-REKA 17
 DB 238 QAMQELHVELERERKA 253

RESULT 4

GCC2_HUMAN GCC2_HUMAN STANDARD; PRT; 1583 AA.
 AC Q8IWT2; O15045; Q8TDS3; Q9H2G8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE GRIP and coiled-coil domain-containing protein 2 (Golgi coiled coil protein GCC185) (CTCL tumor antigen sel-1) (CUL-associated antigen KW-11).
 DE GCC2 OR KIAA0336.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A. (ISOFORM 1).
 RP TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Chira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 [2]

RP SEQUENCE OF 1-448 FROM N.A. (ISOFORM 2).
 RX MEDLINE=22188422; PubMed=12200376;
 RA Krackhardt A.M., Witzens M., Hargis S., Hodi F.S., Zaulis A.J.,
 RA Chessa M., Barrett P., Gribben J.G.;
 RT "Identification of tumor-associated antigens in chronic lymphocytic
 RT leukemia by SEREX";
 RL Blood 100:2123-2131(2002).
 RN [3]
 RN SEQUENCE OF 1-349 FROM N.A. (ISOFORM 1).
 RP TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschuld R.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J.J., Heiton E., Kettunen M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Snailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RN SEQUENCE OF 60-840 FROM N.A. AND TISSUE SPECIFICITY.
 RP TISSUE=Testis;
 RX MEDLINE=21143360; PubMed=11149944;
 RA Eichmüller S., Ueener D., Dummer R., Stein A., Thiel D.,
 RA Schädendorf D.;
 RT "Serological detection of cutaneous T-cell lymphoma-associated
 RT antigens";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
 RN [5]
 RN SUBCELLULAR LOCATION.
 RX MEDLINE=22450606; PubMed=12446665;
 RA Luke M.R., Kjer-Nielsen L., Brown D.L., Stow J.L., Gleeson P.A.;
 RT "GRIP domain-mediated targeting of two new coiled-coil proteins";
 RT GC88 and GC185, to subcompartments of the trans-Golgi network";
 RL J. Biol. Chem. 278:4216-4226(2003).
 CC -!- FUNCTION: Probably involved in maintaining Golgi structure (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; peripheral membrane protein
 CC associated with the trans-Golgi network.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q81WJ2-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q81WJ2-2; Sequence=VSP 007733;
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
 CC -!- SIMILARITY: Belongs to the golgin family.
 CC -!- SIMILARITY: Contains 1 GRIP domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB002334; BAA20794.1; -;
 DR EMBL; AF432211; AAL99918.1; ALT_INIT.
 DR EMBL; BC037774; AAH37774.1; -;

DR EMBL; AF273042; AAG34902.1; ALT_INIT.
 DR Genew; HGNC:23218; GCC2.
 DR InterPro; IPR000237; GRIP domain.
 DR InterPro; IPR003345; M_repeat.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF01465; GRIP; 1.
 DR Pfam; PF02370; N; 9.
 DR PROSITE; PS0913; GRIP; 1.
 KW Golgi stack; Coiled coil; Alternative splicing; Polymorphism.
 FT DOMAIN 9 1517 COILED COIL (POTENTIAL).
 FT DOMAIN 138 947 GLU-RICH.
 FT DOMAIN 1508 1558 GRIP.
 FT VARSPLIC 1 1
 FT VARIANT 1197 1197 R-> G (in dbSNP:1061202).
 FT CONFLICT 593 599 NKKLSSE -> TINSVQK (IN REF. 2).
 FT CONFLICT 836 836 S -> T (IN REF. 2).
 SQ SEQUENCE 1583 AA; 184657 MW; 0FAB4B7EE5969111 CRC64;
 Query Match 50.6%; Score 42; DB 1; Length 1583;
 Best Local Similarity 50.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 QY 2 VQSMRRHRKAELE 15
 DB 325 VQSLKROHQKEISE 338
 |||:::|:|:|:
 RESULT 5
 SPEN DROME STANDARD; PRT; 5560 AA.
 ID SPEN DROME STANDARD; PRT; 5560 AA.
 AC Q8SX83; Q8XHM1; Q8XJ17; Q9UGC3; Q9VPL1; Q9VPL2;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DR Split ends protein.
 GN SPEN OR CGI8497.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Ephydroidea; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC NCBI_TaxID=7227;
 RN [1]_
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4), FUNCTION, ALTERNATIVE PROMOTER
 RP USAGE, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, AND MUTANTS E9 AND
 RP D57.
 RC TISSUE=Embryo;
 RX MEDLINE=20025936; PubMed=10556062;
 RA Kiehllette E.L., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,
 RA McGinnis W.;
 RT "spen encodes an RNP motif protein that interacts with Hox pathways
 RT to repress the development of head-like sclerites in the Drosophila
 RT trunk";
 RL Development 126:5373-5385(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Embryo;
 RX MEDLINE=20157049; PubMed=10655223;
 RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Lavery T.,
 RA Sub C., Voas M., Williams A., Rubin G.M.;
 RT "A genetic screen for novel components of the Ras/mitogen-activated
 RT protein kinase signaling pathway that interact with the yan gene of
 RT Drosophila identifies split ends, a new RNA recognition motif-
 RT containing protein";
 RL Genetics 154:695-712(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Embryo;
 RX MEDLINE=20171275; PubMed=10704397;


```

DR GO: GO:0007411; P:axon guidance; IMP.
DR GO: GO:0008347; P:glia cell migration; IMP.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 3.
DR SMART: SM00360; RRM; 3.
DR PROSITE: PS0102; RRM; 3.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
DR PROSITE: PS0917; SPOC_1.
KW Transcription regulation; Repressor; Developmental protein;
KW Nuclear protein; Repeat; RNA-binding; Coiled coil;
KW Alternative promoter usage; Alternative splicing.
FT DOMAIN 554 632 RNA-BINDING (RRM) 1.
FT DOMAIN 656 730 RNA-BINDING (RRM) 2.
FT DOMAIN 734 806 RNA-BINDING (RRM) 3.
Query Match 50.6%; Score 42; DE 1; Length 5560;
Best Local Similarity 72.7%; Pred. No. 3.7e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 RRRHKAELRK 16
| | | | |
DB 2925 RRRHKAELRK 2935

RESULT 6
YZG1_CABEL STANDARD; PRT; 195 AA.
AC P55325; Q19405;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein F13B6.1 in chromosome X.
GN F13B6.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ernstol N2;
RA Margerison S.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISONS.
RA Durbin R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the TPD52 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z68105; CAA92122.2; -.
DR WormPep; F13B6.1; CE31474.
DR InterPro: IPR007327; TPD52.
DR Pfam: PF04201; TPD52; 1.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 48 97 COILED COIL (POTENTIAL).
SQ SEQUENCE 195 AA; 21287 MW; 5726DE25D519879 CRC64;

Query Match 49.4%; Score 41; DB 1; Length 195;
Best Local Similarity 53.3%; Pred. No. 19;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VQSWRRHKAELRK 16
| | | | |
DB 81 VLSARQKHAELRK 95

```

```

RESULT 7
LECH_RAT
ID LECH_RAT STANDARD; PRT; 283 AA.
AC P02706;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Asialoglycoprotein receptor 1 (Hepatic lectin 1) (RHL-1) (ASGP-R)
DE (ASGPR).
GN ASGRI OR ASGR-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86008335; PubMed=2995379;
RA Leung J.O., Holland E.C., Drickamer K.;
RT "Characterization of the gene encoding the major rat liver
RT asialoglycoprotein receptor.";
RL J. Biol. Chem. 260:12523-12527 (1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85063786; PubMed=6095287;
RA Holland E.C., Leung J.O., Drickamer K.;
RT "Rat liver asialoglycoprotein receptor lacks a cleavable NH2-terminal
RT signal sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:7338-7342 (1984).
RN [3]
RP SEQUENCE OF 11-283 FROM N.A.
RX MEDLINE=87026895; PubMed=2945599;
RA Watts C.;
RT "Isolation and expression of cDNA clones for a rat liver
RT asialoglycoprotein receptor.";
RL Biosci. Rep. 6:527-534 (1986).
CC -1- FUNCTION: Mediates the endocytosis of plasma glycoproteins to
CC which the terminal sialic acid residue on their complex
CC carbohydrate moieties has been removed. The receptor recognizes
CC terminal galactose and N-acetylgalactosamine units. After ligand
CC binding to the receptor, the resulting complex is internalized and
CC transported to a sorting organelle, where receptor and ligand are
CC disassociated. The receptor then returns to the cell membrane
CC surface.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed exclusively in hepatic parenchymal
CC cells.
CC -1- MISCELLANEOUS: Calcium is required for ligand binding.
CC -1- MISCELLANEOUS: TWO TYPES OF RAT HEPATIC LECTIN HAVE BEEN
CC IDENTIFIED, RHL-1 AND RHL-2/3, HAVING A RELATIVE ABUNDANCE OF 4:1.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K02817; AAA42037.1; -.
DR EMBL: M21770; AAA40764.1; -.
DR PIR: A92497; LNRTL.
DR HSP; P20693; IHLJ.
DR InterPro: IPR002353; AntifreezeII.
DR InterPro: IPR001304; Lectin C.
DR InterPro: IPR005640; Lectin N.
DR Pfam: PF00059; lectin C; 1.
DR Pfam: PF03954; lectin N; 1.
DR PRINTS: PRO0356; ANTI-FREEZEII.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C TYPE LECTIN 1; 1.
DR PROSITE: PS0041; C TYPE LECTIN 2; 1.
DR Lectin; Glycoprotein; Receptor; Endocytosis; Transmembrane;
KW

```



```

CC -!- FUNCTION: May participate in forming intercellular cross-bridges
CC of the Golgi complex.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
CC autoimmune disease Sjogren's syndrome.
CC -!- SIMILARITY: Belongs to the golgin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75304; CAA53052.1; -.
CC DR EMBL; D25542; BAA05025.1; -.
CC DR PIR; A56539; A56539.
CC DR PIR; I52300; I52300.
CC DR Genew; HGNC:4429; GOLGB1.
CC DR MIM; 602500; -.
CC DR GO; GO:0000139; C:Golgi membrane; TAS.
CC DR GO; GO:0005795; C:Golgi stack; TAS.
CC DR GO; GO:0016021; C:integral to membrane; TAS.
CC DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
CC KW Golgi stack; Antigen; Coiled coil; Transmembrane.
CC FT DOMAIN 1 3235
CC FT TRANSMEM 3236 3256
CC FT DOMAIN 3257 3259
CC FT DOMAIN 48 593
CC FT DOMAIN 677 1028
CC FT DOMAIN 1062 1245
CC FT DOMAIN 1301 1779
CC FT DOMAIN 1828 3185
CC FT DOMAIN 2420 2423
CC FT DOMAIN 2993 2996
CC FT POLY-SER.
CC FT CONFLICT 1 39 MISSING (IN REF. 3).
CC FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
CC FT CONFLICT 1765 1765 D -> G (IN REF. 3).
CC FT CONFLICT 2950 2950 H -> D (IN REF. 3).
CC SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Query Match 49.4%; Score 41; DB 1; Length 3259;
Best Local Similarity 46.2%; Pred. No. 3.1e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

CY 4 SMREHKAKIREK 16
DB 365 ALEQKHAKREK 377

RESULT 10
TRT DROME
ID TRT DROME STANDARD; PRT; 396 AA.
AC P19351;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DB Troponin T, skeletal muscle (upheld protein) (intended thorax
DE protein).
DE UP OR INT.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Asynchronous muscle;
RX MEDLINE=89141761; PubMed=2852258;
RA Bullard B., Leonard K., Larkins A., Butcher G., Karlik C.,
RA Fyrberg E.A.;

```

```

RT "Troponin of asynchronous flight muscle.";
RL J. Mol. Biol. 204:621-637(1988).
RP REVISIONS, SEQUENCE FROM N.A.
RX STRAIN=Canton-S;
RX MEDLINE=91080155; PubMed=2124273;
RA Fyrberg E.A., Fryberg C.C., Beall C., Saville D.L.;
RT Drosophila melanogaster troponin-T mutations engender three distinct
RT syndromes of myofibrillar abnormalities.;
RL J. Mol. Biol. 216:657-675(1990).
CC -!- FUNCTION: Troponin T is the tropomyosin-binding subunit of
CC troponin, the thin filament regulatory complex which confers
CC calcium-sensitivity to striated muscle actomyosin ATPase activity.
CC Mutations in troponin T engender three distinct syndromes of
CC myofibrillar abnormalities.
CC -!- SIMILARITY: Belongs to the troponin T family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54504; CAA38366.1; -.
CC DR PIR; S13251; S13251.
CC DR Flybase; FBgn004169; up.
CC DR InterPro; IPR001978; Troponin.
CC KW Pfam; PF00992; Troponin; 1.
CC FT DOMAIN 341 396 ASP/GLU-RICH (HIGHLY ACIDIC).
CC SQ SEQUENCE 396 AA; 47333 MW; 3CAA88AA8943C23 CRC64;

Query Match 48.2%; Score 40; DB 1; Length 396;
Best Local Similarity 43.8%; Pred. No. 54;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

CY 2 VQSMREHKAKIREK 17
DB 228 LKELKQKQQLRHKA 243

RESULT 11
TRT SCHPO
ID TRT SCHPO STANDARD; PRT; 477 AA.
AC O13745;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative RNA polymerase II transcription factor TFIIF subunit 1.
GN SPAC16E8.11C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Peltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellton J., Simmonds N., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

```



```

RX MEDLINE=97303186; PubMed=9159129;
RA Radisky D.C., Snyder W.B., Emr S.D., Kaplan J.;
RT "Characterization of VPS41, a gene required for vacuolar trafficking
RT and high-affinity iron transport in yeast.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:5662-5666(1997).
RN [2]
RP SEQUENCE OF 615-744 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95319502; PubMed=7596406;
RA Sherrington R., Rogeev E.I., Liang Y., Rogaeva B.A., Levesque G.,
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
RA Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA Saneau P., Polinsky R.J., Masco W., da Silva H.A.R., Haines J.L.,
RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA Rommens J.M., St George-Hyslop P.H.;
RT "Cloning of a gene bearing missense mutations in early-onset familial
RT Alzheimer's disease.";
RL Nature 375:754-760(1995).
RC -!- FUNCTION: Required for vacuolar assembly and vacuolar traffic.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P49754-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P49754-2; Sequence=VSP_006751; VSP_006752;
CC -!- SIMILARITY: BELONGS TO THE VPS41 FAMILY.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U87309; AAB47563.1; -.
DR EMBL; U87281; AAB47758.1; -.
DR EMBL; L40398; AAC42004.1; -.
DR Genew; HGNC:12713; VPS41.
DR MIM; 605485; -.
DR GO; GO:0016021; C: integral to membrane; TAS.
DR GO; GO:0005478; P: intracellular transporter activity; TAS.
DR GO; GO:0007034; P: vacuolar transport; TAS.
DR InterPro; IPR000547; Clathrin_repeat.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00637; Clathrin; 1.
DR SMART; SM00299; CLH; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR PROSITE; PS00589; ZF_RING_2; 1.
DR Zinc-finger; Alternative splicing.
KW ZNF_RING 791 839 RING-TYPE.
FT DOMAIN 18 27 POLY-GLU.
FT VARSPPLIC 802 802 D -> E (in isoform Short).
FT FTID=VSP_006751.
FT VARSPPLIC 803 854 Missing (in isoform Short).
FT FTID=VSP_006752.
FT CONFLICT 615 618 KQIS -> WHEG (IN REF. 2).
FT CONFLICT 736 744 IPNLRSLV -> DPQFERPLG (IN REF. 2).
SQ SEQUENCE 854 AA; 98535 MW; B30F64D13208550D CRC64;

Query Match 48.2%; Score 40; DB 1; Length 854;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 VQSMRERHKAELRE 15
Db 240 VCSVKERHASEMD 253

RESULT 14

```

```

MYHA BOVIN
ID MYHA BOVIN STANDARD; PRT; 1976 AA.
AC Q27951.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Chaza M., Ishiguro N., Shinagawa M.;
RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 204-302 FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=95301542; PubMed=7782316;
RA Itoh K., Adelstein R.S.;
RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle
RT myosin heavy chain II-B.";
RL J. Biol. Chem. 270:14533-14540(1995).
CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
CC cell shape, and specialized functions such as secretion and
CC capping (by similarity).
CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
CC regulatory light chain subunits (MLC-2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB022023; BAA36494.1; -.
DR EMBL; U15716; AAA87715.1; -.
DR HSPSP; P10587; IBR2.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding; Coiled coil;
KW Multigene family.
FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
FT DOMAIN 786 815 IQ.
FT DOMAIN 845 1976 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP (POTENTIAL).
SQ SEQUENCE 1976 AA; 229097 MW; 6144354451C0F790 CRC64;

Query Match 48.2%; Score 40; DB 1; Length 1976;

```

```

Best Local Similarity 50.0%; Pred. NO. 2.7e+02;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      2 VQSMRERHKAELRE 15
Db      1194 IQDMRQRHATALEE 1207

RESULT 15
MYHA_HUMAN
ID MYHA_HUMAN STANDARD; PRT; 1976 AA.
AC P35580;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
GN MYH10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96025307; PubMed=7499478;
RA Phillips C.L., Yamakawa K., Adelstein R.S.;
RT "Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and
RT analysis of human tissues with isoform-specific antibodies.";
RL J. Muscle Res. Cell Motil. 16:379-389(1995).
RN [2]
RP SEQUENCE OF 63-722 FROM N.A.
RX MEDLINE=91316803; PubMed=1860190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located
RT on different chromosomes.";
RL Circ. Res. 69:530-539(1991).
CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
CC cell shape, and specialized functions such as secretion and
CC capping.
CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
CC regulatory light chain subunits (MLC-2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EXBL; M69181; AAA99177.1; -.
DR PIR; A59252; A59252.
DR HSSP; P10587; 1BR2.
DR Genew; HGNC:7568; MYH10.
DR MIN; 160776; -.
DR GO; GO:0016459; C:myosin; NAS.
DR GO; GO:0003779; F:actin binding; NAS.
DR GO; GO:0005524; F:ATP binding; NAS.
DR GO; GO:0000910; P:cytokinesis; NAS.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.

```

DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding; Coiled coil;
KW Multigene family.
FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
FT DOMAIN 786 815 IQ.
FT DOMAIN 845 1976 COILED COIL (POTENTIAL).
FT NP BIND 178 185 ATP (POTENTIAL).
SQ SEQUENCE 1976 AA; 228938 MW; B2BB87FF35EA124F CRC64;
Query Match 48.2%; Score 40; DB 1; Length 1976;
Best Local Similarity 50.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 2 VQSMRERHKAELRE 15
Db 1194 IQDMRQRHATALEE 1207
Search completed: March 9, 2004, 09:26:03
Job time : 12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 09:24:30 ; Search time 39 Seconds
(without alignments)
137.534 Million cell updates/sec

Title: US-09-998-491-7

Perfect score: 83

Sequence: 1 MVQSMRERHKAELREKA 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 44 | 53.0 | 214 | 16 | Q98GZ7 rhizobium 1 |
| 2 | 44 | 53.0 | 270 | 3 | Q8WZF2 cryptococcu |
| 3 | 44 | 53.0 | 326 | 5 | Q9VW53 drosophila |
| 4 | 44 | 53.0 | 458 | 16 | Q81PE8 bacillus an |
| 5 | 44 | 53.0 | 844 | 3 | Q59891 cryptococcu |
| 6 | 43.5 | 52.4 | 279 | 5 | Q18512 caenorhabdi |
| 7 | 43 | 51.8 | 137 | 2 | Q84CU2 thermus the |
| 8 | 43 | 51.8 | 287 | 4 | Q9ULA6 homo sapien |
| 9 | 43 | 51.8 | 309 | 5 | Q19681 caenorhabdi |
| 10 | 43 | 51.8 | 330 | 11 | Q9CSN1 mus musculus |
| 11 | 43 | 51.8 | 333 | 11 | Q9CV75 mus musculus |
| 12 | 43 | 51.8 | 563 | 5 | Q8MRC3 drosophila |
| 13 | 43 | 51.8 | 720 | 5 | Q9VR99 drosophila |
| 14 | 43 | 51.8 | 743 | 5 | Q9NBV5 drosophila |
| 15 | 43 | 51.8 | 793 | 5 | Q81SC9 plasmodium |
| 16 | 42.5 | 51.2 | 329 | 4 | Q9NUV9 homo sapien |

| | | | | | |
|----|----|------|------|----|--------|
| 17 | 42 | 50.6 | 219 | 2 | Q9LSS4 |
| 18 | 42 | 50.6 | 234 | 16 | Q935Q2 |
| 19 | 42 | 50.6 | 270 | 3 | Q8X232 |
| 20 | 42 | 50.6 | 270 | 3 | Q8WZG1 |
| 21 | 42 | 50.6 | 270 | 3 | Q8WZG9 |
| 22 | 42 | 50.6 | 433 | 16 | Q9X894 |
| 23 | 42 | 50.6 | 703 | 16 | Q83SK1 |
| 24 | 42 | 50.6 | 712 | 11 | Q8BR64 |
| 25 | 42 | 50.6 | 721 | 16 | Q7UDN4 |
| 26 | 42 | 50.6 | 756 | 11 | Q8CSR0 |
| 27 | 42 | 50.6 | 801 | 5 | Q9VK27 |
| 28 | 42 | 50.6 | 801 | 5 | Q8WS36 |
| 29 | 42 | 50.6 | 839 | 5 | Q62002 |
| 30 | 42 | 50.6 | 1029 | 10 | Q80955 |
| 31 | 42 | 50.6 | 1247 | 5 | Q8SRK6 |
| 32 | 42 | 50.6 | 1370 | 13 | Q73732 |
| 33 | 42 | 50.6 | 1379 | 11 | Q62868 |
| 34 | 42 | 50.6 | 1388 | 4 | Q9QUN5 |
| 35 | 42 | 50.6 | 1388 | 4 | Q75116 |
| 36 | 42 | 50.6 | 1388 | 6 | Q28021 |
| 37 | 42 | 50.6 | 1388 | 11 | P70336 |
| 38 | 41 | 49.4 | 201 | 13 | P72M42 |
| 39 | 41 | 49.4 | 246 | 16 | Q8XNR6 |
| 40 | 41 | 49.4 | 420 | 13 | Q8Q3N0 |
| 41 | 41 | 49.4 | 427 | 10 | Q7XL44 |
| 42 | 41 | 49.4 | 503 | 10 | Q7XK33 |
| 43 | 41 | 49.4 | 513 | 10 | Q9LW95 |
| 44 | 41 | 49.4 | 558 | 16 | Q81WT2 |
| 45 | 41 | 49.4 | 565 | 12 | Q8J7B5 |

ALIGNMENTS

RESULT 1
Q98GZ7 ID Q98GZ7 PRELIMINARY; PRT; 214 AA.
AC Q98GZ7;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Transcriptional regulator.
GN MLR3102.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003001; BAB50069.1; -.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001647; HTH_Tetr.
DR Pfam; PF00440; tetr; 1.
DR PRINTS; PR00455; HTHETR.
KW Complete proteome.
SQ SEQUENCE 214 AA; 23794 MW; 755D14AE09F22870 CRC64;
Query Match 53.0%; Score 44; DB 16; Length 214;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 VQSMRERHKAELREK 16

| | | | | | |
|---|---|---|--|--|--|
| DR | Pfam | PF01028; Topoisomerase_I; 1. | | | |
| DR | Pfam | PF02919; Topoisomerase_I_N; 1. | | | |
| DR | PRINTS | PR00416; EUTPTSWRAGEI. | | | |
| DR | SMART | SM00435; TOEUC; 1. | | | |
| KW | Isomerase. | | | | |
| SQ | SEQUENCE | 844 AA; 95995 MW; 3BB32281D1054169 CRC64; | | | |
| | | | | | |
| Query Match | | | | | |
| Best Local Similarity 53.0%; Score 44; DB 3; Length 844; | | | | | |
| Matches 9; Conservative 1; Mismatches 0; Gaps 0; | | | | | |
| | | | | | |
| QY | 7 | ERHKAELREKA 17 | | | |
| | | : | | | |
| DB | 694 | ERHEALELEKA 704 | | | |
| | | | | | |
| RESULT 6 | | | | | |
| Q18512 | PRELIMINARY; PRT; 279 AA. | | | | |
| ID | Q18512 | | | | |
| AC | Q18512; | | | | |
| DT | 01-NOV-1996 (TrEMBLrel. 01, Created) | | | | |
| DT | 01-JAN-1998 (TrEMBLrel. 05, Last sequence update) | | | | |
| DT | 01-JUN-2003 (TrEMBLrel. 24, Last annotation update) | | | | |
| DE | C38D4.7 protein. | | | | |
| GN | C38D4.7. | | | | |
| OS | Caenorhabditis elegans. | | | | |
| OC | Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea; | | | | |
| OC | Rhabditidae; Peloderinae; Caenorhabditis. | | | | |
| OX | NCBI_TaxID=6239; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | Coles L.; | | | | |
| RL | Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases. | | | | |
| [2] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | none; | | | | |
| RX | MEDLINE=99069613; PubMed=9851916; | | | | |
| RT | "Genome sequence of the nematode C.elegans: A platform for | | | | |
| RT | investigating biology."; | | | | |
| RL | Science 282:2012-2018 (1998). | | | | |
| DR | EMBL; Z46241; CAA86321.1; -. | | | | |
| DR | PIR; T19828; T19828. | | | | |
| DR | WormPep; C38D4.7; CE00920. | | | | |
| DR | InterPro; IPR007087; Znf C2H2. | | | | |
| DR | SMART; SM00355; Znf C2H2. 1. | | | | |
| DR | PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1. | | | | |
| KW | Metal-binding; zinc; zinc-finger. | | | | |
| SQ | SEQUENCE 279 AA; 30809 MW; 221845F75B821BEC CRC64; | | | | |
| | | | | | |
| Query Match | | | | | |
| Best Local Similarity 66.7%; Pred. No. 51; | | | | | |
| Matches 10; Conservative 3; Mismatches 1; Indels 1; Gaps 1; | | | | | |
| | | | | | |
| QY | 3 | QSMERHKAELREKA 17 | | | |
| | | : | | | |
| DB | 110 | RSIREHRAE-REKA 123 | | | |
| | | | | | |
| RESULT 7 | | | | | |
| Q84CU2 | PRELIMINARY; PRT; 137 AA. | | | | |
| ID | Q84CU2 | | | | |
| AC | Q84CU2; | | | | |
| DT | 01-JUN-2003 (TrEMBLrel. 24, Created) | | | | |
| DT | 01-JUN-2003 (TrEMBLrel. 24, Last sequence update) | | | | |
| DT | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) | | | | |
| DE | Nucleoside diphosphate kinase. | | | | |
| GN | NDX. | | | | |
| OS | Thermus thermophilus. | | | | |
| OC | Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae; | | | | |
| OC | Thermus. | | | | |
| OX | NCBI_TaxID=274; | | | | |
| [1] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | |

```

RC STRAIN=HB8;
RA Takeishi S., Nakagawa M., Maoka N., Kihara M., Moriguchi M., Masui R.,
RA Kuramitsu S.;
RT "Crystallization and preliminary X-ray diffraction studies of
RT nucleoside diphosphate kinase from Thermus thermophilus HB8.";
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB107688; BAC7699.1; -.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0016301; P:kinase activity; IEA.
DR GO; GO:0004550; P:nucleoside-diphosphate kinase activity; IEA.
DR GO; GO:0006241; P:CTP biosynthesis; IEA.
DR GO; GO:0006183; P:GTP biosynthesis; IEA.
DR GO; GO:0006228; P:UTP biosynthesis; IEA.
DR InterPro; IPR001564; NDK.
DR Pfam; PF00334; NDK; 1.
DR PRINTS; PR01243; NUCDPKINASE.
DR PRODOM; PD001018; NDK; 1.
DR SMART; SM00562; NDK; 1.
KW Kinase.
SQ SEQUENCE 137 AA; 15344 MW; 750E673CA9C9FE3E CRC64;

Query Match 51.8%; Score 43; DB 2; Length 137;
Best Local Similarity 64.3%; Pred. No. 30;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 QSMRERHKAELREK 16
DB 42 QELAEHYAEHREK 55

RESULT 8
Q9ULA6 PRELIMINARY; PRT; 287 AA.
AC Q9ULA6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Nuclear receptor coactivator NCOA-62 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Baradarani L., Birditt B.,
RA Bloom S., Dors M., Dickhoff R., Harrison G., James R., Lasky S.,
RA Madan A., Ratcliffe A., Shaffer T., Hood L.;
RT "Sequencing of human chromosome 14.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC008044; AAF01479.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0004872; P:receptor activity; IEA.
DR InterPro; IPR004015; SKIP_SNW.
DR Pfam; PF02731; SKIP_SNW; 1.
KW Receptor.
FT NON TER
SQ SEQUENCE 287 AA; 33463 MW; 07CB28612609414B CRC64;

Query Match 51.8%; Score 43; DB 4; Length 287;
Best Local Similarity 47.1%; Pred. No. 63;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 MVQSMRERHKAELREKA 17
DB 54 MAQKEKEKEKLEKRWMA 70

RESULT 9
Q19681 PRELIMINARY; PRT; 309 AA.
AC Q19681;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

```

```

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F21D5.2 protein.
GN F21D5.2
OS Caenorhabditis elegans.
OC Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Berks M.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z54271; CAA91033.1; -.
DR PIR; T21195; T21195.
DR WormPep; F21D5.2; CE03245.
DR InterPro; IPR003323; OTU.
DR Pfam; PF02338; OTU; 1.
DR PROSITE; PS08002; OTU; 1.
SQ SEQUENCE 309 AA; 34704 MW; 2A15EC6CC12F18F3 CRC64;

Query Match 51.8%; Score 43; DB 5; Length 309;
Best Local Similarity 60.0%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MVQSMRERHKAELRE 15
DB 53 MEKEMDSRHKKELEE 67

RESULT 10
Q9CSN1 PRELIMINARY; PRT; 330 AA.
AC Q9CSN1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 11 days embryo whole body cDNA, RIKEN full-length enriched library,
DE clone:2700046D12 product:nuclear protein SKIP (SKI-interacting
DE protein) (SNW1 protein) (Nuclear receptor coactivator NCOA-62)
DE homolog.
GN SKIP.
OS Mus musculus (Mouse).
OC Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carrinci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanaoki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Komoto H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Soqabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```

| | |
|---|--|
| RA | Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., |
| RA | Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., |
| RA | Saito T., Okazaki Y., Gotojori T., Bono H., Kasekawa T., Saiko R., |
| RA | Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavanti T., |
| RA | Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., |
| RA | Kuehl P., Lewis S., Matsuo Y., Nikaïdo I., Pesole G., Quackenbush J., |
| RA | Schriml L.M., Stauber P., Suzuki R., Tomita M., Wagner L., Washio T., |
| RA | Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., |
| RA | Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., |
| RA | Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., |
| RA | Gunstinch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., |
| RA | Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., |
| RA | Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., |
| RA | Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P., |
| RA | Suzuki H., Toyok-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L., |
| RA | Yamash-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseuki S., |
| RA | Hayashizaki Y.; |
| RT | "Functional annotation of a full-length mouse cDNA collection."; |
| RT | Nature 409:685-690(2001). |
| DR | EMBL; AK009218; BAB26144.1; -- |
| DR | MGB; MGI:1913604; Sklip. |
| DR | GO; GO:0005634; C:nucleus; IEA. |
| DR | InterPro; IPR004015; SKIP SNW. |
| DR | Pfam; PF02731; SKIP SNW; I. |
| FT | NON TER 1 |
| SQ | SEQUENCE 333 AA; 38772 MW; 7CEE761694815672 CRC64; |
| Query Match 37.58%; Score 43; DB 11; Length 333; | |
| Best Local Similarity 47.1%; Pred. No. 72; | |
| Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps | |
| QY | 1 MYQSMRERHKAIREKA 17 |
| DB | 109 MAQKQKREHEKUREMA 125 |
| RESULT 12 | |
| Q8WRC3 | |
| ID | Q8WRC3 PRELIMINARY; PRT; 563 AA. |
| AC | Q8WRC3 |
| DT | 01-OCT-2002 (TrEMBLrel. 22, Created) |
| DT | 01-OCT-2002 (TrEMBLrel. 22, Last sequence update) |
| DT | 01-OCT-2002 (TrEMBLrel. 25, Last annotation update) |
| DE | RE148580. |
| GN | CACTIN OR CG1676. |
| OS | Drosophila melanogaster (Fruit fly). |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; |
| OC | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; |
| OC | Ephydroidea; Drosophilidae; Drosophila. |
| CC | NCBI TaxID=7227; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=Berkley; |
| RA | Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J., |
| RA | Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E., |
| RA | George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G., |
| RA | Miranda A., Mungall C.J., Nunoo J., Pacble J., Paragas V., Park S., |
| RA | Patel S., Phouanavong S., Wan X., Yu C., Lewis S.E., Rubin G.M., |
| RA | Celniker S.; |
| RL | Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases. |
| DR | EMBL; AY121670; AAN51957.1; -- |
| DR | FlyBase; FBgn0031114; cactin. |
| DR | GO; GO:000950; P:dorsal/ventral axis specification; IGI. |
| SQ | SEQUENCE 563 AA; 67271 MW; FBC382C3FB27FDD0 CRC64; |
| Query Match 51.8%; Score 43; DB 5; Length 563; | |
| Best Local Similarity 66.%; Pred. No. 1.2e+02; | |
| Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps | |
| QY | 5 MRERHKAIREK 16 |
| DB | 273 LRERHQAIREK 284 |

Query Match 51.8%; Score 43; DB 5; Length 720;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Wed Mar 10 09:07:50 2004

Search completed: March 9, 2004, 09:26:56
Job time : 41 secs

us-09-998-491-7.rspt